Combining Molecular Dynamics Simulations and Atomic Force Microscopy Experiments to Rationalize the Mechanical Properties of Double-Stranded DNA and RNA

Universidad Autónoma de Madrid
Departamento de Física Teórica
de la Materia Condensada

Dissertation submitted to obtain the degree of *Doctor en Física*

Author: Alberto Marín González
Supervisors: Prof. Fernando Moreno-Herrero / Prof. Rubén Pérez Pérez

Madrid, February 2020

Acknowledgements

As a recipient of an international PhD fellowship from "La Caixa Severo Ochoa" program, I would like to acknowledge the support received all these years from "La Caixa Foundation".

I would also like to acknowledge the computer resources and technical expertise provided by the Spanish Supercomputer Network (RES) at Minotauro Supercomputer (BSC, Barcelona) and Altamira (IFCA, Cantabria) and by the computing facilities of the Extremadura Research Centre for Advanced Technologies-Center for Energy, Environment, and Technology (CETA-CIEMAT).

Contents

Abstract

Many biological processes interrogate the mechanical properties of double-stranded DNA (dsDNA) and RNA (dsRNA). Such processes rely on the ability for dsDNA and dsRNA to deform when subjected to mechanical stress or upon interaction with other molecules. From the local motion of individual base pairs upon protein binding, to the global folding of genome-long polymers, these distortions span a wide range of length scales. While immense efforts have been devoted to unveil the mechanical properties of these duplexes, connecting the dynamics at such multiple length scales still remains a challenging problem.

In this PhD dissertation, we present a holistic approach to this problem. We adopt a multiscale perspective, where the level of spatial resolution is thoroughly varied according to the specific question at hand. To that end, we combine atomistic molecular dynamics (MD) simulations, single-molecule experiments, and theoretical models. MD simulations allow us to explore the dynamics of the duplexes at the atomic scale and make predictions on how these motions are translated into global, polymer-like mechanical properties. By means of single-molecule techniques we experimentally test some of our computational predictions and measure novel effects that might motivate future theoretical developments. Lastly, simple theoretical models are built to rationalize and bring together the simulation results and the experimental observations.

According to the strategy employed, this Thesis is divided in two parts. In the first part (Chapter 2) we follow a *top-down* approach. Previous single-molecule measurements performed on long DNA and RNA duplexes motivated us to scrutinize atomistic mechanisms by means of MD simulations. Of particular importance here is the opposite torsional response reported for dsDNA

and dsRNA under an external force: dsDNA overwinds when stretched whereas dsRNA unwinds. Guided by this experimental observation, we revealed an opposite change of the dsDNA and dsRNA interstrand distance upon elongation.

In the second part (Chapters 3 to 6), the strategy followed is predominantly *bottom-up*: molecular features of particular dsDNA and dsRNA sequences are theoretically predicted (Chapters 3 and 4) or experimentally measured (Chapters 5 and 6) to give rise to certain macroscopic phenomena. In Chapters 3 and 4, we use MD simulations to study the mechanical response of different dsDNA and dsRNA sequences to an external force. In Chapter 3, we unveil that the sequence-dependent dsDNA stretching flexibility is encoded in the shape of the molecule via a structural feature that we name *crookedness*. In Chapter 4, we report that the nucleotide sequence affects in a strikingly divergent manner the mechanical response of dsRNA and dsDNA. We find that this effect is a consequence of the large flexibility of dsRNA pyrimidine-purine steps. In chapters 5 and 6 we dwell on two peculiar kinds of sequences: dsDNA A-tracts and dsRNA AU-tracts. The first kind is known to introduce a bend in the DNA and has been well characterized at the local level, while its global mechanical properties have remained controversial. In Chapter 5, we build upon the well-known local features of A-tracts to provide a comprehensive description of their global flexibility, disentangling the different mechanical properties of these sequences by means of a combination of single-molecule techniques. Motivated by the dsDNA A-tracts, we report in Chapter 6 a systematic study of sequence-induced bending in dsRNA. Firstly, MD simulations are used to identify a sequence motif, which we name AU-tract, which stabilizes bent conformations in dsRNA. This computational prediction is then supported by atomic force microscopy (AFM) imaging experiments, which reveal that dsRNA molecules rich in AU-tracts are more prone to adopt bent conformations than control dsRNA molecules of arbitrary sequence.

Through the examples presented, this Thesis highlights how the combination of MD simulations and single-molecule techniques can contribute towards bridging the gap between the different length scales involved in the mechanical properties of dsDNA and dsRNA.

Resumen

Las propiedades mecánicas del ADN y el ARN de doble hélice (dsDNA y dsRNA, por sus siglas en inglés) juegan un papel clave en un gran número de procesos biológicos. En estos procesos, las moléculas de dsDNA y dsRNA se deforman debido a estrés mecánico o a la interacción con otras moléculas. Desde el movimiento local de pares de bases al interaccionar con proteínas hasta el plegamiento de genomas enteros, estas distorsiones abarcan diversas escalas de longitud. Por ello, para entender cuantitativamente las funciones de estos ácidos nucleicos, es preciso caracterizar sus propiedades mecánicas desde una perspectiva multiescala.

En esta Tesis doctoral presentamos un estudio multiescala sobre las propiedades mecánicas del dsDNA y el dsRNA, donde el nivel de resolución espacial se ajusta a responder la pregunta en cuestión. Para ello, combinamos simulaciones de dinámica molecular (MD), experimentos de molécula individual y modelos teóricos. Las simulaciones de MD nos permiten explorar la dinámica de las moléculas de dsDNA y dsRNA a nivel atómico, así como hacer predicciones sobre cómo estos movimientos se traducen en propiedades mecánicas de polímeros a nivel global. Usando técnicas de molécula individual, comprobamos experimentalmente algunas de nuestras predicciones computacionales y medimos nuevos fenómenos que podrían motivar futuros desarrollos teóricos. Finalmente, desarrollamos modelos teóricos simples para racionalizar e integrar los resultados de las simulaciones y las observaciones experimentales.

En función de la estrategia utilizada, esta Tesis está dividida en dos partes. La primera parte (Capítulo 2) representa un enfoque *top-down*. Usando simulaciones de MD, logramos reproducir los parámetros mecánicos de las moléculas

de dsDNA y dsRNA obtenidos previamente en experimentos de molécula individual. Cabe destacar aquí la respuesta torsional opuesta del dsDNA y dsRNA sometidos a una fuerza externa, según la cual el dsDNA se enrolla al estirarse mientras que el dsRNA se desenrolla. Guiados por esta observación experimental, indagamos en la dinámica de las moléculas a nivel atomísitico y revelamos que, al ser estiradas, la distancia entre las hebras del dsDNA y del dsRNA siguen una tendencia opuesta. En la segunda parte (Capítulos del 3 al 6), la estrategia seguida es principalmente *bottom-up*: predecimos teóricamente (Capítulos 3 y 4) y medimos experimentalmente (Capítulos 5 y 6) cómo ciertas características moleculares de determinadas secuencias de dsDNA y dsRNA dan lugar a efectos macroscópicos. En los Capítulos 5 y 6, usamos simulaciones de MD para estudiar la respuesta mecánica de diferentes secuencias de dsDNA y dsRNA sometidas a una fuerza externa. En el Capítulo 3 desvelamos que la flexibilidad de estiramiento del dsDNA está codificado en la forma de la molécula, por medio de un mecanismo estructural que denominamos *crookedness*. En el Capítulo 4, mostramos que la secuencia de nucleótidos afecta de una forma completamente diferente la respuesta mecánica del dsRNA y el dsDNA. Este efecto se puede entender gracias a la gran flexibilidad de los dinucleótidos pirimidina-purina del dsRNA. Los Capítulos 5 y 6 tratan sobre dos tipos de secuencias: los *A-tracts* del dsDNA y los *AU-tracts* del dsRNA. Los primeros son conocidos por doblar la molécula de dsDNA a nivel local, sin embargo sus propiedades mecánicas globales no han sido completamente caracterizadas. En el Capítulo 5, partiendo de las características locales de los *A-tracts*, desarrollamos una descripción de su flexibilidad global, desentrañando las diferentes propiedades mecánicas de estas secuencias mediante una combinación de varias técnicas de molécula individual. En el Capítulo 6, motivados por el caso de los *A-tracts* del dsDNA, describimos un estudio sistemático sobre cómo la secuencia dobla la molécula de dsRNA. En primer lugar, usamos simulaciones de MD para identificar una secuencia patrón, que llamamos *AU-tract*, que estabiliza conformaciones dobladas del dsRNA. Después, mediante experimentos de AFM desvelamos que las moléculas de dsRNA ricas en *AU-tracts* son más propensas a doblarse comparadas con moléculas de secuencias arbitraria, apoyando las predicciones computacionales.

En esta Tesis demostramos el potencial de combinar simulaciones de MD y experimentos de molécula individual de cara a comprender los complejos mecanismos multiescala que rigen las propiedades mecánicas del DNA y el RNA de doble hebra.

1 Introduction

1.1 A historical perspective on Nucleic Acids

Deoxyribonucleic (DNA) and ribonucleic (RNA) acids, collectively termed nucleic acids (NA), are central to all known forms of life. Possibly, the first evidence that NA are an essential component of living organisms can be attributed to Avery, MacLeod and McCarty, who observed that DNA (and not protein) is the substance responsible for bacterial transformation, and thus, the carrier of genetic information [1]. This idea gained gradual acceptance in the following years, most notably in 1952, when Hershey and Chase found that, during bacteriophage infection, DNA is transferred from the virus to the host [2]. Finally, the role of DNA as the genetic material became evident in 1953 from the structure of the molecule. Based on previous X-ray diffraction data [3, 4] and base complementarity rules [5], Watson and Crick proposed their double helical structure of DNA, which immediately suggested a mechanism for copying and transferring genetic information [6]. Since then, DNA has occupied a preeminent position in research, attracting scientists from disciplines that range from physics to medicine, including paleontology or criminology.

The unique role of DNA as storage of genetic information is in stark contrast with the plethora of functions that are performed by RNA. Indeed, as of 2019, 30 scientists have been awarded a Nobel Prize for their experimental work on RNA. The classical role of RNA as a messenger in the translation of the DNA genetic information into proteins was postulated in the late 1950's by Francis Crick and was experimentally demonstrated in the 1960's, when the efforts of Severo Ochoa and Marshall Nirenberg, among others, succeeded in deciphering

the genetic code [7–10]. A decade later, it was found that this messenger RNA (mRNA) is often split prior to translation, in a process known as RNA splicing. The versatile character of RNA was further confirmed with the discovery of ribozymes, RNA molecules with the capability of catalyzing chemical reactions. In the last decades, RNA has received renewed attention, prompted, in part, by newly discovered functions of RNA in its double helical form (dsRNA). In fact, dsRNA has been shown to mediate gene silencing [11], trigger the immune response against viral attacks [12, 13] or constitute a crucial component of large ribonucleoprotein complexes such as the ribosome [14, 15].

1.2 Basics of Nucleic Acids

1.2.1 The chemical composition of DNA and RNA

Nucleic acids are polymers, or chains formed by the assembly of repeating subunits, whose building blocks are nucleotides. These consist of a pentose sugar attached to a phosphate group and a nitrogenous base, or nucleobase (Fig. 1.1a). The latter is the most essential component of NA, since a chain of bases can be linearly arranged in endless combinations, or sequences, and thus provide a molecular mechanism for storing information. There are four canonical DNA bases: thymine, cytosine, adenine and guanine. The first two contain a single six-membered ring and are known as pyrimidines; the other two comprise an additional five-membered ring fused to the pyrimidine ring and are classified as purines. The standard RNA bases are the same as in DNA, only that thymine is substituted by uracil, which lacks the methyl group at the C5 carbon. In addition, non-standard DNA and RNA bases are usually found in nature, which often arise by adding a chemical group to the canonical bases. The most representative example is 5-methylcytosine, which is highly abundant in the genomes of several species.

When attached to a pentose sugar, the nucleobase forms a nucleoside. In DNA the sugar is 2-deoxy-D-ribose whereas in RNA it is D-ribose, the difference being the presence of an additional –OH group in the 2' carbon of the latter. Then, a nucleoside can be joined to a phosphate group through the 5' carbon

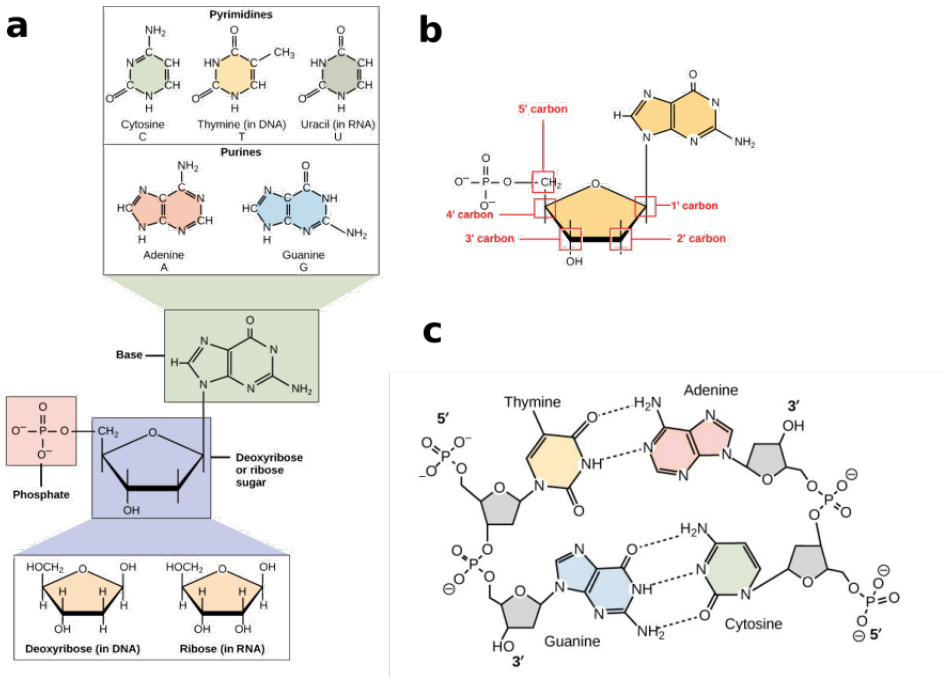

Figure 1.1: Basic chemical features of nucleic acids. (a) Representation of the three components of a nucleotide: the base (green), the sugar (blue) and the phosphate group (pink). The different kinds of DNA and RNA bases and sugars are shown. (b) A nucleotide with the atoms of the sugar marked. (c) Representation of Watson-Crick base pairs in a DNA dimer. Image adapted from *www.khanacademy.com*.

of the sugar, constituting a nucleotide (Fig. 1.1b). Finally, a dinucleotide is built by attaching the phosphate group of a given nucleotide to the 3' carbon of the sugar of a second nucleotide via a phosphodiester bond. By iterating this process, a nucleic acid chain is generated, which can be seen as a sugar-phosphate backbone to which the nucleobases are joined. The extremes of this chain are denoted by 3'-end and 5'-end, in allusion to the carbon atoms of the sugar proximal to that end (Fig. 1.1c).

1.2.2 The structure of DNA and RNA double-helices from fiber diffraction

Virtually all the biological processes that involve nucleic acids, from replication to transcription or translation, rely on the ability of the nucleobases to form base pairs by means of hydrogen bonding interactions. The most common type is called Watson-Crick base pairing and refers to the interaction between adenine and thymine/uracil and between guanine and cytosine, which are stabilized by two and three hydrogen bonds respectively. When two strands or two fragments of a given strand can engage in complete Watson-Crick base pairing while running antiparallel to each other (that is one strand in the 3' to 5' and the other in the 5' to 3' direction), they are said to be complementary. In this case, base pairing results in the formation of a double helix, which is further stabilized by the stacking interactions between adjacent base pairs.

Interestingly, the double helical structure formed by double-stranded DNA (dsDNA) presents remarkable differences with the structure adopted by double-stranded RNA (dsRNA), see Fig. 1.2. DsDNA commonly adopts a B-form conformation, which is a right-handed helix with a helical pitch of 3.4 nm, 10.5 bp per turn and a radius of 1 nm. In B-DNA, the base pairs are slightly displaced ($\sim$0.8 Å) from the helical axis and are almost completely perpendicular to it, presenting a low inclination of $\sim$2.4 deg. Another important structural feature of this double-helix concerns the grooves, which are the voids between the strands. In particular, there is a strong asymmetry between the size of the grooves, being the major groove significantly wider ($\sim$22 Å) than the minor groove ($\sim$12 Å). On the other hand, dsRNA displays a double-helical conformation known as

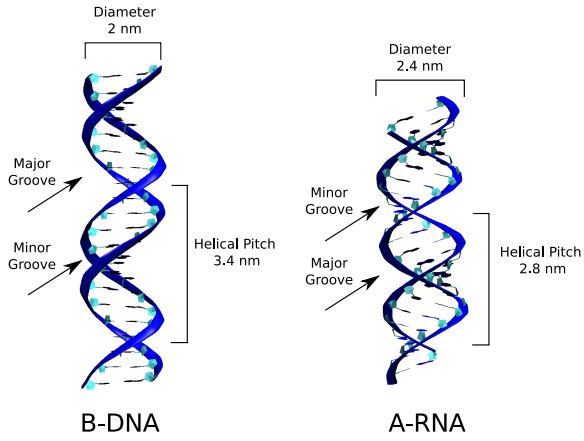

Figure 1.2: DNA and RNA double-helices. Images were generated using the VMD software [18].

A-form. This is also a right-handed helix, but it is shorter and wider than the B-form, with a helical pitch of ∼2.8 nm, 11 bp per turn and a radius of ∼1.2 nm [16]. In addition, the base pairs are strongly inclined (∼12 deg) and displaced (∼4.1 Å) with respect to the helical axis [17]. Finally, the size of the grooves is quite different in the A- and B-forms. Contrary to the B-form, in the A-form the major and the minor grooves have a similar width.

1.2.3 Sequence effects in the structures of NA duplexes

The model B- and A-forms captured the general aspects of the early dsDNA and dsRNA structures obtained from fiber diffraction experiments. Nevertheless, with the advent of single-crystal structures and the associated boost in resolution, it soon became clear that the dsDNA structure is not uniform, but presents local sequence-dependent variations [19]. These sequence effects arise as a combination of three factors: a) the unfavourable steric repulsion between atoms in close proximity; b) the attractive $\pi - \pi$ interaction between adjacent base pairs; and c) the constrains imposed by the backbone chain [10, 20–22]. Sequence effects are also expected in the dsRNA structure, but an experimental approach to this issue remains challenging, mostly due to the scarce structural data available for dsRNA [23]. Importantly, sequence-dependent structural fea-

tures are key to the biology of nucleic acids, since they provide a mechanism for proteins to recognize a particular NA sequence.

In general, sequence-dependent dsDNA/dsRNA conformations consist on local displacements and rotations between adjacent base pairs (see Fig. 1.2). In some cases, these local features are amplified, and can greatly affect the global dimensions of the helix. Well-known examples of this are the poly-G and poly-A DNA sequences. The former stabilize a structure that resembles an A-form helix [24], whereas the latter adopt a non-canonical conformation that induces a bend in the DNA [25, 26].

In order to characterize the sequence-dependent conformation of a given dsDNA/dsRNA molecule, three sets of parameters are often used, that we will collectively denote as rigid body parameters, see Fig. 1.3. The first set describes the geometry of a given base pair and is made of three translational (Shear, Stretch and Stagger) and three rotational parameters (Buckle, Propeller and Opening). Analogously, the set of base pair step parameters represents the relative position (Shift, Slide and Rise) and orientation (Tilt, Roll and Twist) between two adjacent base pairs. These two sets of parameters provide a complete description of the conformations adopted by each base pair in a given NA molecule. Indeed, by knowing all the base pair and base pair step parameters of a particular sequence, it is possible to reconstitute the atomic structure of the NA [27]. Finally, a third set of parameters quantifies the geometry of a given base pair with respect to the helical axis. Of particular importance are the helical rise and helical twist parameters. These parameters quantify the extension and the number of turns per bp, which are the typical quantities measured in single-molecule experiments such as magnetic or optical tweezers. The other four parameters represent the displacement (x and y) of the base pair from the helical axis and the angles between the helical axis and the plane defined by the base pair.

1.3 Mechanical characterization of nucleic acids

The structures of dsDNA and dsRNA are insufficient to explain many of the functions of these molecules inside the cell. Rather than their structure, many

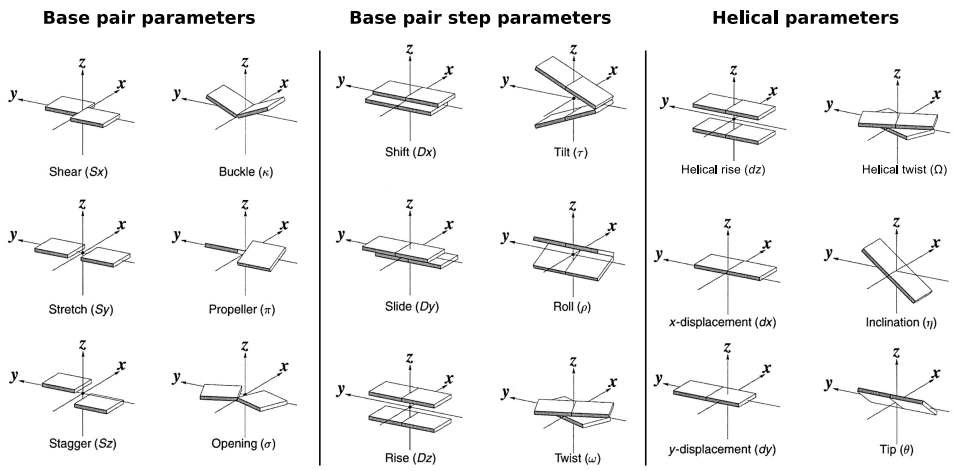

Figure 1.3: Schematic representation of the rigid body parameters. Figure adapted from *x3dna.org*.

biological processes interrogate the flexibility of these molecules, that is, their ability to adopt a conformation that deviates from their relaxed equilibrium state. Therefore, in order to quantitatively understand these processes, a comprehensive description of the mechanical properties of NA is required.

A key aspect of the mechanical properties of dsDNA and dsRNA is that they are highly multi-scale, spanning distances that range from single base pairs (Å) to several kilo base pairs (μm). At the level of individual base pairs, local changes in the NA conformation are exploited by several proteins to achieve specific sequence recognition via an indirect readout mechanism. At a global scale, NA chains are bent, twisted and stretched in processes such as the packaging of genomic dsRNA inside the viral capsid, the formation of DNA loops at regulatory regions or the compaction of DNA into nucleosomes. As a consequence of this multi-scale character, numerous descriptions and models of NA flexibility have been proposed that account for varying levels of detail.

1.3.1 A global approach: single-molecule experiments

In the last decades, the invention and continuous improvement of single-molecule techniques have yielded an unprecedented quantitative picture of the global me-

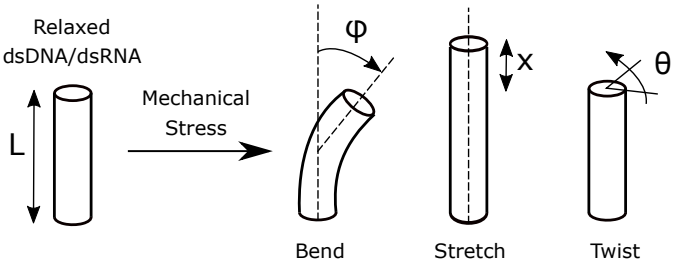

Figure 1.4: Cartoon depicting the deformations of an elastic rod under mechanical stress. This model is usually employed to describe the mechanical properties of dsDNA and dsRNA in single-molecule experiments.

chanical properties of dsDNA and dsRNA [28]. In single-molecule experiments, NA duplexes are usually modelled as elastic rods that can undergo bending, stretching and twisting deformations, see Fig. 1.4. In this model, the energetic cost associated to deform the NA from its relaxed structure is assumed to be quadratic and can be expressed as [1]

$$E = \frac{1}{2}\frac{B}{L}\varphi^2 + \frac{1}{2}\frac{S}{L}x^2 + \frac{1}{2}\frac{C}{L}\theta^2 + \frac{g}{L}x\theta, \tag{1.1}$$

where B is the bending rigidity, S the stretch modulus, C the twist rigidity and g the twist-stretch coupling. The quantities L, φ, x, and θ are depicted in Fig. 1.4 and represent respectively the extension of the molecule in its relaxed state and the bending angle, elongation and twisting angle induced by mechanical stress. In the following we will briefly describe how the four mechanical parameters - B, S, C and g - have been determined for dsDNA and dsRNA in a series of benchmark single-molecule experiments.

[1] Two disclaimers are to be made with respect to Eq. 1.1. Firstly, we are neglecting the coupling between bending and twisting deformations, whose value was recently measured for dsDNA [29] but is still missing in the dsRNA case. Secondly, the assumption that bending deformations are harmonic is currently a subject of debate. For example, an AFM work proposed that, at short length scales, this relation should be linear instead [30].

Bending rigidity of dsDNA and dsRNA

Obtaining B from single-molecule measurements requires taking into account the entropic forces that dominate the dynamics of the long (>100 bp) NA molecules typically used in these experiments. This is usually done by resorting to the worm-like chain (WLC) model, which describes the bending deformations of polymers in the framework of statistical mechanics. In this model, the bendability of the polymer is described by its persistence length, P, which is related to its bending rigidity according to $B = k_B T P$.

The persistence length of dsDNA and dsRNA has been measured by depositing these molecules on a mica surface and imaging them using an atomic force microscope (AFM). If the molecules are properly adsorbed, they will adopt 2D equilibrated conformations that can be fit to the 2D WLC model, thus allowing to obtain P [31]. These AFM measurements have yielded values of $P_{DNA} =$ 51 $\pm$ 1 nm and $P_{RNA} = 60 \pm 1$ nm for dsDNA and dsRNA respectively [32].

The values of P for dsDNA and dsRNA can also be extracted from force spectroscopy experiments using magnetic (MT) and optical tweezers (OT). In the MT set-up, a controlled force can be exerted on a dsDNA or dsRNA molecule that is tethered between a magnetic bead and a surface. The extension of the molecule is tracked by measuring the diffraction patterns of the magnetic bead using an inverted microscope. In an OT set-up, a dsDNA/dsRNA molecule is tethered between two polystyrene beads. These beads are held in optical traps, which allow to move the beads with high precision, thus permitting to stretch the NA molecule in a highly controlled manner [33]. The force is then obtained from the displacement of the beads from the center of their traps. In the range of forces lower than $\sim$10 pN, the characteristic MT and OT force-extension curves of dsDNA and dsRNA molecules are well described by the WLC model [34, 35]. The values of the dsDNA and dsRNA persistence length thus obtained were P_{DNA} $\sim$50 nm and P_{RNA} $\sim$60 nm, in good agreement with the AFM measurements [32].

The stretch modulus

At forces larger than ∼10 pN, dsDNA and dsRNA molecules are almost fully extended to their contour length and their stretching elasticity can be probed by OT [32, 34, 36] and, more recently, also MT [37]. The stretching flexibility has been successfully incorporated into the WLC resulting in the so-called extensible WLC (eWLC) model, that can be used to fit force-extension (Fz) curves in the $F < 40$ pN range [34, 36]. Using this approach, Herrero-Galan *et al.* computed the dsDNA and dsRNA stretch moduli from OT data, revealing a substantial difference between both NA duplexes: S_{DNA} =935±120 pN, S_{RNA} =500±29 pN [32]. The softer stretching flexibility of dsRNA compared to dsDNA was later confirmed by MT experiments, which suggested an even larger difference (around three fold) between their stretch moduli [37].

The twist rigidity

A direct determination of the twist rigidity of the NA duplexes requires an experimental technique capable of measuring both torques and forces [28]. This has been achieved by developing new experimental methods, including the rotor bead assay in OT [38] and variations of the MT set-up, such as magnetic torque tweezers (MTT) [39, 40] or freely-orbiting magnetic tweezers (FOMT) [41]. In a recent work, MTT have been used to measure the twist rigidity of dsDNA and dsRNA, reporting similar values for both molecules: C_{DNA} ∼450 pN nm^2 and C_{RNA} ∼410 pN nm^2 [37].

Twist-stretch coupling

Intuitively, one would expect that a double-helix will unwind upon stretching. The reason is that when the turns of the helix are removed, the two parallel strands reach a longer extension, see Fig. 1.5. However, this is not the case for DNA at low forces (<20 pN). In fact, fine measurements of the changes in the twisting angle as an external force was exerted on the duplex revealed that DNA overwinds when stretched [42]. Consistently, when the molecule was overtwisted by means of MT, its extension increased [42, 43]. Since then, additional

Figure 1.5: DNA wood toy illustrating the unwinding-when-stretched behaviour intuitively expected for a double-helix. Contrary to this simple picture, forces < 20 pN result in overwinding of DNA [42]. Image adapted from *www.cornerstonescareer.com*.

measurements of the DNA twist-stretch coupling have confirmed this effect, employing, for instance, gold rotor bead tracking [44] or incorporating twisting deformations into the WLC model [45, 46].

Interestingly, the opposite behaviour was reported for dsRNA, for which application of positive turns resulted in a decrease of the extension [37]. This opposite trend in the mechanical response of dsDNA and dsRNA is expressed mathematically by the opposite sign of their twist-stretch coupling parameter: $g_{DNA}/k_BT \sim -20$, $g_{RNA}/k_BT \sim$11.5 [37, 42].

1.3.2 A local approach: analysis of structures from NA databases

A major drawback of single-molecule measurements is that they lack information on how dsDNA and dsRNA respond to mechanical stress at the local level of a few base pairs, where DNA-protein interactions usually take place. A description of these local effects could elucidate how the different nucleobases affect the mechanical properties of the duplexes, shedding light on how the sequence might regulate biological processes that rely on the physical properties of these molecules.

The local deformability of NA can be assessed by means of experimental techniques, such as nuclear magnetic resonance (NMR) or X-ray crystallography,

that resolve the structure of NA duplexes. By analysing extensive data sets of NA structures, the base pair flexibility of the duplexes can be inferred, as pioneered by Olson et al [47]. In that work, a systematic description of sequence-dependent DNA deformability is provided in the context of the base pair step parameters (see Fig. 1.3) using a simple mechanical model based on two assumptions. The first one considers that the deformations of a base pair step are harmonic in the coordinate system of the base pair step parameters. The second assumption is the dinucleotide or nearest neighbor approximation, according to which the flexibility of a given base pair step is only dependent on the sequence comprising that particular dinucleotide. The elastic energy of each dinucleotide can then be written as

$$E = \frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} k_{ij}\Delta q_i \Delta q_j, \tag{1.2}$$

where the indices i, j run over the six base pair step parameters; $k_{i,j}$ are the elastic constants associated to each deformation; and $\Delta q_i = q_i - q_i^0$ is the deviation of the i^{th} parameter from its value at the minimum energy configuration. Note that the resulting stiffness matrix $(K)_{ij} \equiv k_{i,j}$ is not diagonal, such that the different deformations can be coupled, as occurred in the elastic rod model (Eq. 1.1).

Using this approach, Olson et al revealed novel features of the dynamics of DNA, such as the enhanced flexibility of pyrimidine-purine dinucleotides or the importance of the off-diagonal terms of the stiffness matrix [47]. However, the rigid body models derived from the analysis of structural data sets present significant shortcomings. For instance, a certain sequence of interest can be under-represented or completely absent from the data-base. This issue is crucial in the case of dsRNA, for which the number of structures available is very limited and often present artifacts [23]. More importantly, models derived from structural data sometimes fail to capture the global features of NA flexibility described in Section 1.3.1. Concretely, attempts to determine the elastic rod parameters of dsDNA and dsRNA from the analysis of crystal structures yielded paradoxical results and were incapable of reproducing, for example, the opposite twist-stretch coupling of the duplexes [48].

1.3.3 A versatile tool: Molecular Dynamics

All-atom molecular dynamics (MD) simulations are an excellent tool to circumvent these problems. Triggered by the continuous refinements of the NA force fields and the steady increase in computational power, MD can now provide descriptions of NA flexibility at different levels of detail, in quantitative agreement with the experimental observations.

From a local perspective, MD have been successfully applied to reveal fine details of the dynamics of dsDNA. Of particular relevance are the studies performed by an international collaboration of MD groups, the ABC consortium, which have extensively characterized the sequence-dependent conformational landscape of dsDNA, rationalizing the limitations of the harmonic and the nearest-neighbor approximations [49–52]. In parallel, a large number of MD works have addressed the global mechanical properties of dsDNA, determining the stretching [53–65], twisting [57–69] and bending [53–64, 70, 71] flexibility of different sequences as well as the twist-stretch coupling [56, 63, 65, 72, 73]. The local and global flexibility of dsRNA have also been investigated, although the literature is more scarce than in the dsDNA case [23, 58, 61, 65, 71, 72, 74, 75].

Nevertheless, despite the huge success of MD simulations in determining the mechanical properties of NA duplexes at different resolution levels, a crucial question remains largely unanswered, namely, how the dynamics at these different length scales are interconnected.

1.4 Outline of this Thesis

In this Thesis, we combine MD simulations and single-molecule experiments to pursue an understanding of the mechanical properties of dsDNA and dsRNA from a multi-length scale perspective. The gain of this approach is two-fold. On one hand, the macroscopic properties of dsDNA and dsRNA measured in single-molecule experiments motivate the exploration of novel motions at the atomic scale. This *top-down* strategy is presented in Chapter 2, where previous experimental measurements of the dsDNA and dsRNA twist-stretch coupling prompted our analysis of sub-Å changes in the interstrand distance of the du-

plexes. On the other hand, the molecular features of certain dsDNA and dsRNA sequences are linked with particular mechanical properties that can be measured in single-molecule experiments. This is the approach followed in Chapters 3 to 6. The complete path that connects molecular effects to macroscopic mechanical properties is best illustrated in Chapter 6. In that Chapter, we show how sub-nm changes in the size of the major groove of certain dsRNA sequences can affect the flexibility of ∼200 nm-long duplexes, as quantified by changes in their persistence length.

In order to establish the connection between MD simulations and single-molecule experiments, we resort to the elastic rod model presented in Section 1.3.1, which reasonably describes the dynamics of the duplexes over all the length-scales considered. Through this Thesis, we focus our attention on three of the elastic rod mechanical parameters: the stretch modulus, the bending rigidity or persistence length, and the twist-stretch coupling. In Chapter 2, where we deal with the mechanical properties of arbitrary dsDNA and dsRNA sequences, the relevant deformations are the stretching and the twist-stretch coupling. The reason is that these two deformations are substantially different for dsDNA and dsRNA (see Section 1.3.1), and thus, they guide our search for qualitative differences in the dynamics of the duplexes at the atomic-scale. In the rest of the Thesis, we study the stretching and bending deformations because they are more easily accessible to standard single-molecule techniques. Hence, stretching and bending distortions provide a more direct route towards experimental validation of our computational and theoretical results. More concretely, in Chapters 3 and 4 we are mainly concerned with the stretching stiffness of several dsDNA and dsRNA molecules. Finally, in Chapters 5 and 6, we mostly deal with the bending deformations of particular dsDNA and dsRNA sequences, which we probe experimentally using AFM imaging.

1.5 Methods

In the following, we will briefly describe the methods employed here to measure the stretching, twist-stretch coupling and bending deformations of dsDNA and dsRNA at different length scales.

1.5.1 Constant-Force Molecular Dynamics

The stretch modulus of dsDNA and dsRNA have been experimentally determined by exerting a controlled force on the duplexes and measuring the corresponding change in extension [34]. Similarly, the dsDNA twist-stretch coupling was obtained by tracking the changes in twisting angle upon the action of an external force [42]. Inspired by these experiments, we developed a protocol to apply a controlled force on a NA duplex during the course of an MD simulation. As will be explained in Chapter 2, these constant-force molecular dynamics (CFMD) simulations were able to neatly reproduce both the opposite twist-stretch coupling of the two NA duplexes and the three-fold softer dsRNA stretching response. Importantly, this was achieved from average measurements, *i.e.* without the need to rely on the fluctuations of the systems. Nevertheless, these fluctuations are required to fully determine the elastic rod model parameters.

Exerting a constant force on NA duplexes

The force was exerted on the base pairs adjacent to the terminal base pairs of the duplexes, see Fig. 1.6a. To that end, we defined virtual atoms (purple beads in Fig. 1.6a) that are located at the middle point between the C1' atoms (red beads) of the next-to-terminal base pairs. These virtual atoms do not represent real, physical atoms, but are also subjected to the equations of motion and hence, can be easily incorporated into an MD simulation. The external force acts on these virtual atoms, which redistribute the load to the C1' atoms. Note that, to prevent from possible fraying effects, we avoided stretching from the terminal base pairs. In addition, we chose atoms that do not belong to the nucleobases, but to the phosphate backbone, so as to minimize the effect of the force on the base pairing interactions.

A natural way of introducing mechanical stress in an MD simulation is by defining a scalar potential, commonly known as restraining potential, $U(R)$. The variable R is the restraining coordinate, which in our case is the distance between the virtual atoms. Then, the associated force is equal to the negative

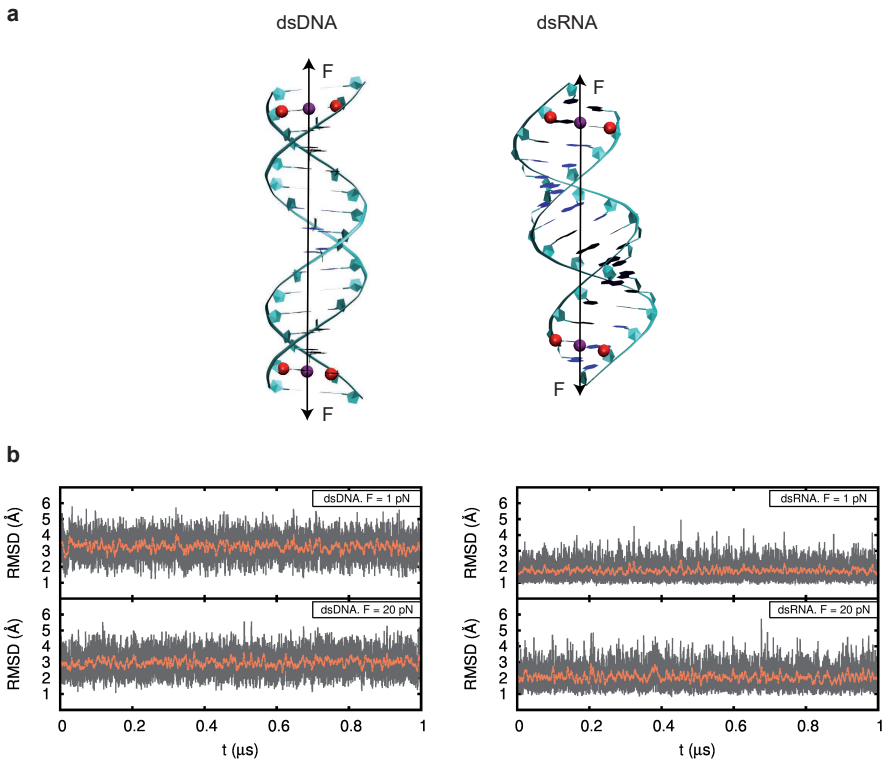

Figure 1.6: DsDNA and dsRNA molecules under a constant stretching force. (a) The force was implemented to act on the centers (purple beads) of the C1' atoms of the second and next-to-last base pairs (red). (b) Root mean squared deviations (RMSD) of the heavy atoms of the 10 central base pairs of 16 bp-long dsDNA and dsRNA molecules were computed with respect to their standard B- and A-forms. Gray: RMSD at every simulation frame (2 fs); orange: RMSD averaged over a running window of 2,000 frames.

gradient of the restraining potential, that is,

$$\vec{F} = -\vec{\nabla}U(R). \tag{1.3}$$

Due to the widespread use of steered MD, most simulation packages incorporate harmonic restraining potentials, which yield a force that is linear with respect to R. However, we are interested in introducing a constant force, which according to Eq. 1.3 requires a linear restraining potential. We achieved this by resorting to a piecewise-defined function that is implemented in the latest versions of the *AMBER* software package [76]. The shape of this restraining potential is depicted in Fig. 1.7 and presents a linear region at values of $R < r_1$, where r_1 is an input parameter defined by the user. Because the slope in that region is negative, Eq. 1.3 states that the resulting force will be positive in the direction of increasing R, and thus, a stretching force. Therefore, by choosing an arbitrarily large value of r_1, one can exert a constant stretching force on the duplex. The magnitude of the force can be accurately fixed by tuning other input parameters in the *AMBER* software. Further technical details of the CFMD implementation are presented in Appendix A.

General Simulation Details

In our CFMD simulations, we initially built the NA duplexes using the software NAB [76] and we then neutralized the molecules with sodium counter-ions. No additional salt was added. The systems were then placed in a cubic box of explicit water molecules, with size dimensions of ∼100 Å and were subjected to energy minimization in two stages: a first stage of 5000 steps with restrains on the NA and a second stage of 5000 steps without any restrain on the system. The systems were then heated up to a temperature of T =300 K and were equilibrated for 20 ns at constant pressure and temperature (NPT ensemble). After this NPT equilibration, production CFMD simulations were run starting from the last configuration of the NPT equilibration. These CFMD runs were performed at constant volume and temperature (NVT ensemble) and were extended to 1 μs of simulation time.

In this Thesis, CFMD simulations were performed using the *AMBER 14*

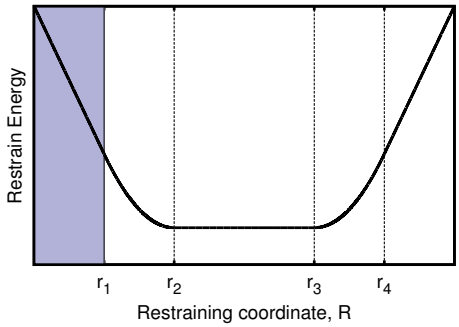

Figure 1.7: Shape of the restraining potential recently implemented in the AMBER software package [76]. The parameters r_1, r_2, r_3 and r_4 split the space of the restraining coordinate in five regions. If the restraining coordinate remains in the first region (blue) the potential is linear with negative slope, resulting in a constant stretching force acting on the atoms.

software package [76] with NVIDIA GPU acceleration [77–79]. We used the parmbs0 [80] and the χ_{OL3} [81] modifications of the Cornell ff99 force field [82] to describe dsDNA and dsRNA; Joung/Cheatham parameters for the sodium counter-ions [83]; and the TIP3P model for water [84]. To account for electrostatic interactions, we used periodic boundary conditions and Particle Mesh Ewald (with default values and a real-space cutoff of 9 Å). Van der Waals interactions were truncated at the real-space cutoff. Bonds containing hydrogen were constrained using the SHAKE algorithm. We used an integration step of 2 fs and saved the coordinates every 1000 steps.

The MD trajectories were analized using Ambertools [76]. We resorted to the 3DNA [27] convention to obtain the rigid body parameters, unless otherwise stated. Helical rise and helical twist parameters were used to measure the overall extension and twisting angle of the NA duplexes, as done elsewhere [72, 85]. Groove dimensions were measured using Curves+ [86]. Errors were computed by splitting the trajectories in five time windows and computing the SEM of the measurements in these subtrajectories. The base pairs located close to the termini of the duplexes were not included in the analysis in order to avoid end-

effects. We thus discarded three base pairs at each termini (*i.e.* a total of six) of the duplex in Chapter 2 and four in the rest of the Thesis.

In this Thesis, MD simulations were performed and analyzed with the help of Dr. J. Guilherme Vilhena.

Initial Checks

As a first characterization, we performed 1 μs-long CFMD simulations of 16 bp dsDNA and dsRNA molecules. The studied sequence was the same for dsDNA and dsRNA, only changing T by U: 5'-GCGCAATGGAGTACGC. We simulated the duplexes under different values of the stretching force and found that these simulations are stable in the 1-20 pN force range. This can be seen from the RMSD of the duplexes, which are similar at external forces of 1 and 20 pN, see Fig. 1.6b. A more demanding test is presented in Chapter 2, where we demonstrate that our CFMD simulations reproduce the linear force-extension and force-twist behaviour predicted by the elastic rod model, yielding values of the elastic parameters that are in good agreement with single-molecule experiments.

1.5.2 Atomic Force Microscopy Imaging

In AFM imaging, the topology of a sample surface can be characterized by scanning the surface with an oscillating tip and measuring the amplitude, frequency and phase shift of the oscillation [87]. Using an AFM, one can image dsDNA and dsRNA molecules adsorbed on a surface and, under proper conditions, the spatial trajectories of these polymers will report on their bending flexibility [32].

Here, dsDNA and dsRNA molecules were imaged using an AFM from Nanotec Electronica S.L. with PointProbePlus tips (PPP-NCH Nanosensors). Images were taken using amplitude modulation (tapping mode) in air, at room temperature and humidity. Sample images were acquired at a resolution of 1-2 nm per pixel and processed using the WSxM software [88].

A precise characterization of the persistence length of dsDNA and dsRNA from AFM imaging requires that the molecules adopt 2D equilibrated conformations when adsorbed on the surface. We achieved this by resorting to previously reported protocols that have been shown to enable proper equilibration of dsDNA

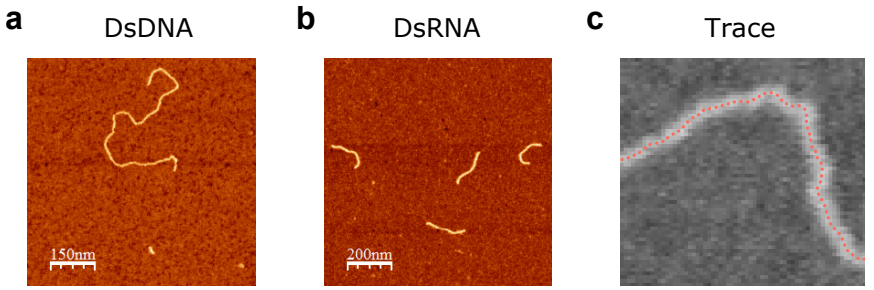

Figure 1.8: Representative AFM images of a dsDNA molecule (a) and dsRNA molecules (b) adsorbed using the protocol described in this section. (c) Fragment of a typical trace computed using the routine published in [30].

and dsRNA molecules on a mica surface [32, 89]. We further added 100 mM NaCl to the reported mixtures, in order to account for the Na^+ ions that we use in the MD simulations (see Section 1.5.1). For dsDNA, we prepared a 20 μL solution containing 0.3 nM of DNA in 100 mM NaCl, 10 mM Tris-HCl pH 8 and 15 mM $MgCl_2$. DsRNA molecules were incubated in a solution of 0.5 nM dsRNA, 2.5 mM $NiCl_2$, 2.5 mM Tris-Ac pH 7.5, 2.5 mM MgOAc and 100 mM NaCl. These solutions were deposited onto freshly cleaved mica and incubated for 30-60 s. The sample was then washed with Milli-Q water and dried using air nitrogen.

Representative AFM images of the dsDNA and dsRNA molecules studied in this Thesis are shown in Fig. 1.8a, b. The contour lengths of the molecules can be directly measured from the images using the WSxM software [88], previous application of an image flattening filter. In order to assess the mechanical properties of the molecules, the images need to be converted into traces, that is, 2D curves that follow the contours of the polymers. This was achieved by resorting to a LabView tracing routine written by Fernando Moreno-Herrero and published in [30]. Based on previous works, we chose a point-to-point separation of l=2.5 nm [30, 89]. An example trace is shown in Fig. 1.8c.

AFM imaging and analysis was done together with Alejandro Martin-Gonzalez (Chapter 5 and 6) and Mikel Marin-Baquero (Chapter 6) from Fernando Moreno-Herrero's lab.

2 General Mechanical Properties of dsDNA and dsRNA Derived from MD Simulations

Adapted from: A. Marin-Gonzalez, J. G. Vilhena, R. Perez, F. Moreno-Herrero *Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics*, Proc. Natl. Acad. Sci. USA **114**, 7049 (2017).

2.1 Introduction

In the past decades, a particular effort has been devoted to unveil the complex mechanisms that drive the mechanical response of nucleic acids (NA) [32, 37, 42, 43, 90, 91]. They occur in nature in two chemically different molecules: DNA and RNA, both possessing the ability to form complementary double helices. In-vivo, NA are not typically found in their relaxed forms. Instead, in the biological processes they take part, proteins wrap [92], bend [93], stretch [94] and twist [95] double stranded DNA (dsDNA) and RNA (dsRNA) molecules [96]. As a result of these mechanical stresses, the NA structure and functionality are affected [16]. A direct evidence of the interplay between mechanical-stress and biological-activity has been provided via single-molecule experiments [97, 98]. Indeed, techniques such as optical (OT) and magnetic tweezers (MT) [90, 97, 98] have emerged as essential tools in the characterization of the mechanical properties of dsDNA [90], and most recently dsRNA [32, 37, 99].

Single molecule experiments are performed on very long NA chains (thousands of base-pairs), thus making extremely challenging to access how the mechanical stress changes the NA structure locally (i.e., how it changes at the scale

at which proteins operate). This is especially important for the small forces (below 20 pN) at which most biological processes occur [35]. The analysis of single-molecule force-extension data with a continuum mechanics approach, the so-called elastic rod model [100, 101], reveals two particularly striking results (see Section 1.3.1): the three-fold softer stretching response of dsRNA compared to dsDNA [32], and the opposite sign of the twist-stretch coupling [37, 42, 43, 72]. Ad-hoc model extensions, like the addition of an outer helical wire [42] and several models at the base-pair level [37, 48] have been proposed to rationalize some of these results. However, no satisfactory explanation for the twist-stretch coupling difference between dsDNA and dsRNA at the continuum or coarse grain level has been found. A satisfactory explanation may involve differences in the atomic-scale structure that are beyond a continuum mechanics analysis [28, 72].

To investigate local NA structural changes induced by mechanical stress one may resort to all atom molecular dynamics (MD) simulations. MD can access the atomic detail of a NA structure [102–105] and provide information about the mechanical response of the chain as a whole [64, 72, 103, 106–108]. One common approach to address this problem is to determine the elastic parameters of a given NA chain through thermal fluctuations at the equilibrium of an unrestrained sequence [56, 66, 72]. However, this approach cannot explore extensions corresponding to large forces regimes. This results in the incapability of understanding why dsDNA twist-stretch coupling changes with an applied force [42, 43, 45, 46]. An alternative approach computes potentials of mean force (PMF) [106, 108, 109]. Nevertheless, obtaining a PMF, where a given reaction coordinate (*e.g.*, elongation or twist) is sampled enough to ensure thermal equilibrium is reached at each point, is a very demanding computational task. As a result, in most MD simulations reported so far, the PMF is computed from many short simulations (t <100 ns). Although they provide very useful insights about conformational changes arising from deformation, the statistical fluctuations of the observables is too large to provide an accurate quantitative prediction of NA mechanical parameters [108, 110].

Here we overcome these limitations and provide a comprehensive characterization of the mechanical response of both dsDNA and dsRNA based on μs-long

all-atom MD simulations of these molecules under constant stretching forces in the 1-20 pN range. From the response of the average elongation, the average twist, and the coupling of their fluctuations to the applied force, we determine all the elastic constants of dsDNA and dsRNA without any prior assumption. Our data agree with general elastic values reported in the literature and reproduce the twist-stretch coupling force-dependence measured in single-molecule experiments. A thorough analysis of our simulations provides a firm basis to explain the differences between both molecules. Firstly, we develop the springiness hypothesis into a discrete model to understand the three-fold softer stretching response of dsRNA. Secondly, we explain the opposite sign in the twist-stretch coupling of dsDNA and dsRNA by changes in their inter-strand distance. Analysis of our data at the atomistic level allowed us to trace down this different behavior and the complex force dependence of the twist-stretch coupling to the response of the slide base pair parameter, and its relation with the sugar pucker angle.

2.2 Measuring the mechanical parameters of NA

We performed CFMD simulations using the 16-mer dsDNA and dsRNA molecules and the simulation details described in Section 1.5.1. Our simulations were first analysed with the elastic rod model [100, 101] (see Section 1.3.1). As bending fluctuations can be neglected on the length scale of the molecule [72], the energy (E) of a stretched and twisted NA molecule subjected to a constant force (F), may be written as [42]

$$E = \frac{1}{2}\frac{S}{L}x^2 + \frac{1}{2}\frac{C}{L}\theta^2 + \frac{g}{L}x\theta - xF \tag{2.1}$$

where L is the equilibrium extension at zero force, x is the elongation or deviation from L, and θ is the change in helical twist from its unperturbed equilibrium value. As defined in Section 1.3.1, the three constants S, C, and g are the stretch modulus, the twist rigidity, and the twist-stretch coupling, respectively. From Eq. 2.1, one can derive the elastic rod model constitutive equations for a NA molecule (see Appendix B):

$$\frac{x}{L} = \frac{1}{\tilde{S}} F, \tag{2.2}$$

$$\frac{\theta}{L} = -\frac{g}{C\tilde{S}} F, \tag{2.3}$$

$$\frac{\partial x}{\partial \theta} = -\frac{g}{S} \tag{2.4}$$

where $\tilde{S} = S - g^2/C$ is the effective stretch modulus and values are obtained at equilibrium. Therefore, to determine the NA elastic parameters S, C, and g, it suffices to measure the force dependence of three observables, *i.e.*, the average elongation $x = f(F)$, the average twist deformation $\theta = f(F)$, and how thermal fluctuations of the twist are coupled to fluctuations of the elongation, $\partial x/\partial\theta$.

2.2.1 Determination of NA elastic parameters from MD simulations

As predicted by the elastic rod model (Eq. 2.2, 2.3), x and θ changed linearly with the applied force (Fig. 2.1a, b), thus corroborating that the elongation and twist deformation are elastic in the range of forces employed here. The slope of a linear fit to the extension data provided the twist unrestrained stretch modulus $\tilde{S}$ of the dsDNA and dsRNA molecules (Fig. 2.1a). For both molecules, $\tilde{S}$ was positive and the values provided by our simulations, $\tilde{S}_{DNA}$ = 1120 $\pm$ 50 pN and $\tilde{S}_{RNA}$ = 416 $\pm$ 7 pN were consistent with experimental results (Table 2.1) and reproduced the three-fold softer behavior of dsRNA previously experimentally measured [32].

The torsional mechanical response of dsDNA and dsRNA (Fig. 2.1b) showed that as the force increases DNA overwinds while RNA unwinds. Although counter-intuitive, this behavior is in agreement with previous experimental and theoretical works [37, 42, 43, 72]. The slope of a linear fit to the helical twist data provided a measurement of $-g/(C\tilde{S})$ (Fig. 2.1b). Given that both $\tilde{S}$ and C are positive, a negative slope leads to a positive value of g, *i.e.*, unwinding of dsRNA, and a positive slope to a negative value of g, *i.e.*, overwinding of dsDNA. DsDNA and dsRNA showed twist rigidities of C_{DNA} = 303 $\pm$ 23 pN nm^2 and C_{RNA} = 310 $\pm$ 24 pN nm^2, in good agreement with experimental measurements (Table 2.1).

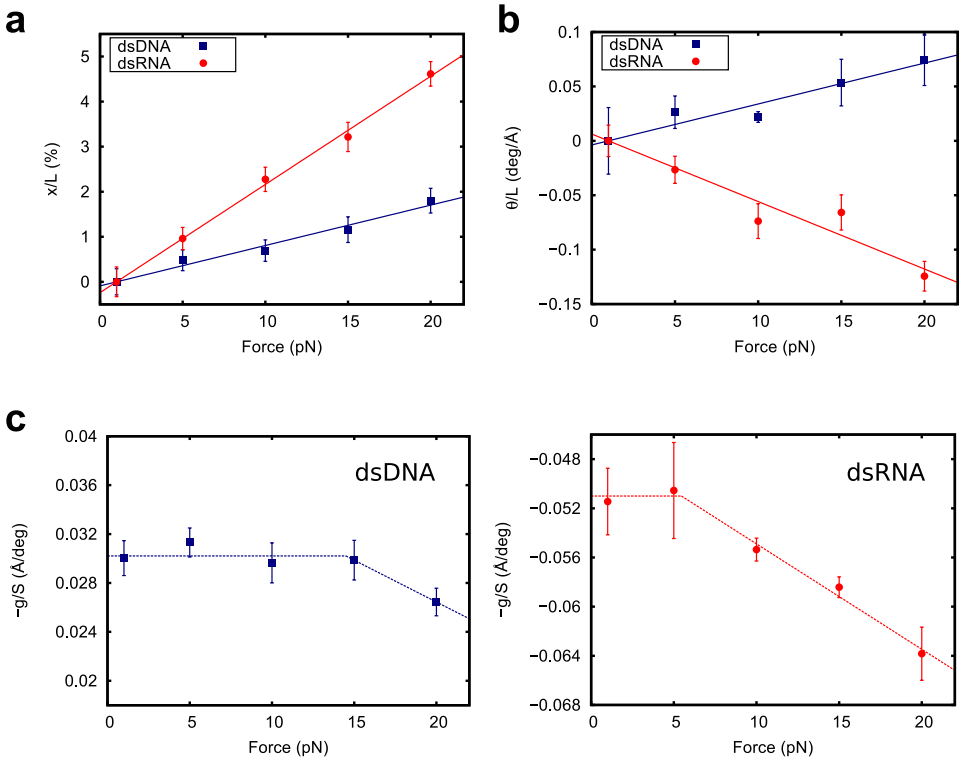

Figure 2.1: Obtaining the elastic rod model parameters from MD data. (a) Elongation of dsDNA (blue) and dsRNA (red) divided by the extension at $F = 1$ pN plotted as a function of the applied force. The extension was computed as the mean value of the helical rises of the 10 central bp averaged over the last 0.8 μs (400,000 simulation frames) at each constant force. The linear fits have slopes of $(89 \pm 4)\times10^{-5}$ pN^{-1} and $(240 \pm 4)\times10^{-5}$ pN^{-1} for dsDNA and dsRNA, respectively. (b) Change in the twisting angle of dsDNA and dsRNA with respect to the simulation data at $F = 1$ pN divided by the extension at $F = 1$ pN, plotted as a function of the force. The twisting angle was computed as the mean of the helical twists of the 10 central bp averaged over the last 0.8 μs. A linear fit was performed, yielding slopes of $(3.8 \pm 0.3)\times10^{-3}$ deg Å^{-1} pN^{-1} and $(-6.2 \pm 0.6)\times10^{-3}$ deg Å^{-1} pN^{-1} for dsDNA and dsRNA, respectively. (c) The ratio $-g/S$ was computed at each constant force simulation as the slope of the linear fit of the helical twist as a function of the helical rise (see Fig. 2.2). Lines are a guide to the eye. Linear fits in (a) and (b) were constrained to pass through the point (1,0). Error bars were calculated by splitting the data into five time windows of 160 ns and calculating the SEM considering the measurements in each window.

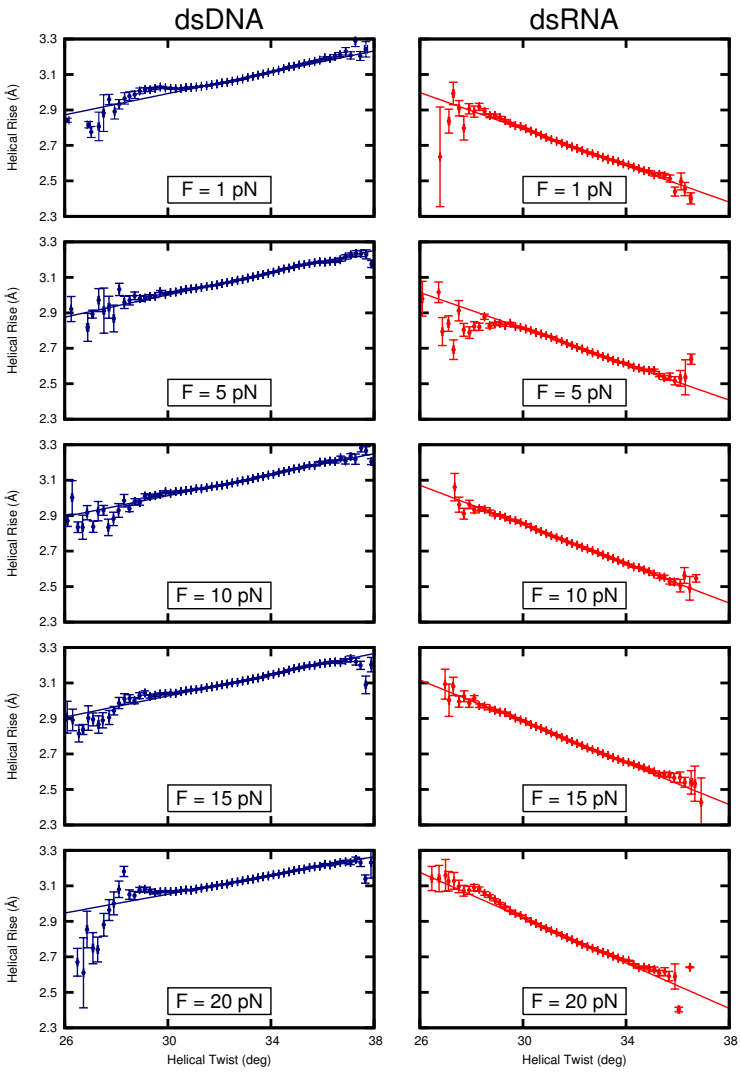

Figure 2.2: Helical rise - helical twist slopes at different forces for dsDNA/dsRNA. The helical twist and helical rise per base pair were computed for each simulation frame. The slopes were computed by fitting this raw data. For representation purposes and following Ref. [72], the helical twist was discretized in bins of width of 0.2 degrees and the mean helical rise and helical twist was computed in each bin. Error bars are the SEM in each bin. The slopes of the fits at the different forces are shown in Fig. 2.1.

Mining each of our ten different μs MD trajectories following the methods reported in Ref. [72], we computed the correlations in helical twist and helical rise fluctuations at different forces (Fig. 2.2). The values of these correlations are shown in Fig. 2.1c and report on how the twist-stretch coupling changes with the external force. Our data leads to two remarkable observations. Firstly, g has a negative value for dsDNA and a positive value for dsRNA, providing an independent confirmation of the opposite twist-stretch coupling of these molecules. Although these values are in reasonable agreement with a previous theoretical work [72], they showed an overestimation in absolute values compared with experimental measurements (Table 2.1, see below for further discussion). Secondly, the twist-stretch coupling for dsDNA and dsRNA changed with force. While this dependence affected dsDNA for forces beyond 15 pN, in agreement with recent experimental results [45], changes in g^{RNA} started at lower forces (5-10 pN). For instance, a change in force of 20 pN implied an increase of g^{RNA} of $\sim$20%, *i.e.*, it unwound more easily when stretched. The only experimental measurement of g^{RNA} has been reported by Lipfert el al. [37] in the range of forces between 4 and 8 pN, where they found g to be force-independent. According to our simulations, in this range one should observe a small change of $\sim$4%, which is clearly below the resolution of the reported experiments.

In the following, we will take advantage of the atomistic detail of our simulations to dwell on the origin of three-fold softer stretching response of dsRNA compared to dsDNA, the opposite sign in their twist-stretch coupling, and its dependence with force.

2.3 Explaining the different dsDNA and dsRNA stretch modulus

We developed a discrete model based on the *springiness* hypothesis proposed in [37, 48]. Our springiness model is explained in detail in Appendix C, here we will simply outline its main features. In short, our model considers a polymer that joins the centers of all consecutive base pairs, as shown in Fig. 2.3a. If bending is neglected, the extension of the molecule can be written as

$$h = h(l, \cos\beta) = l\cos\beta \tag{2.5}$$

Table 2.1: The constitutive equations of the elastic rod model (Eq. 2.2-2.4) allowed us to compute all elastic parameters from the slopes of Fig. 2.1a, b and from the helical rise-helical twist slopes (Fig. 2.1c). Quoted errors represent deviations from the linear fits. Constants obtained in this work are compared with other experimental (exp.) works.

Elastic Parameter	DNA (This work)	DNA (exp.)	RNA (This work)	RNA (exp.)
S (pN)	1260 ± 110	1450-1750 [45, 46]	480 ± 21	
$\tilde{S}$ (pN)	1100 ± 90	649-1401 [32, 34, 36, 37, 111]	418 ± 13	350,500 [32, 37]
C (pN nm^2)	300 ± 40	386-448 [37, 38, 112, 113]	320 ± 60	409 [37]
g/k_BT (unitless)	-54 ± 5	-17 to -39 [37, 42, 44–46, 112]	34 ± 1	11.5 [37]

where l is the contour length of this centerline polymer and β is the angle defined by this polymer with the helical axis (z-axis in Fig. 2.3b), thus providing a measurement of the *springiness* of the system. Here we propose that an increase in extension can be achieved either by decreasing β or increasing l. Therefore, the elongation (x) is, at first order approximation, the sum of these two contributions:

$$x \equiv \Delta h = x_{\Delta\beta} + x_{\Delta l}. \tag{2.6}$$

We calculated the $x_{\Delta\beta}$ and $x_{\Delta l}$ contributions for dsDNA and dsRNA for different forces and obtained a linear behavior (Fig. 2.3c), supporting the assumption of the springiness model (Eq. C.7). Then, the inverse of the slopes of linear fits to the data sets $x(F)/L$, $x_{\Delta\beta}(F)/L$, and $x_{\Delta l}(F)/L$ are respectively the elastic constants $\tilde{S}$, k_β, and k_l. (Fig. 2.3c and Eq. B.6, C.7). Note that $\tilde{S}$ can be also computed from k_β and k_l as (Eq. C.9)

$$\tilde{S} = \frac{k_\beta k_l}{k_\beta + k_l} \tag{2.7}$$

We computed the value of $\tilde{S}$ from this equation and obtained $\tilde{S}_{DNA} = 1076 \pm 68$ pN and $\tilde{S}_{RNA} = 421 \pm 15$ pN. These values are in good agreement with the

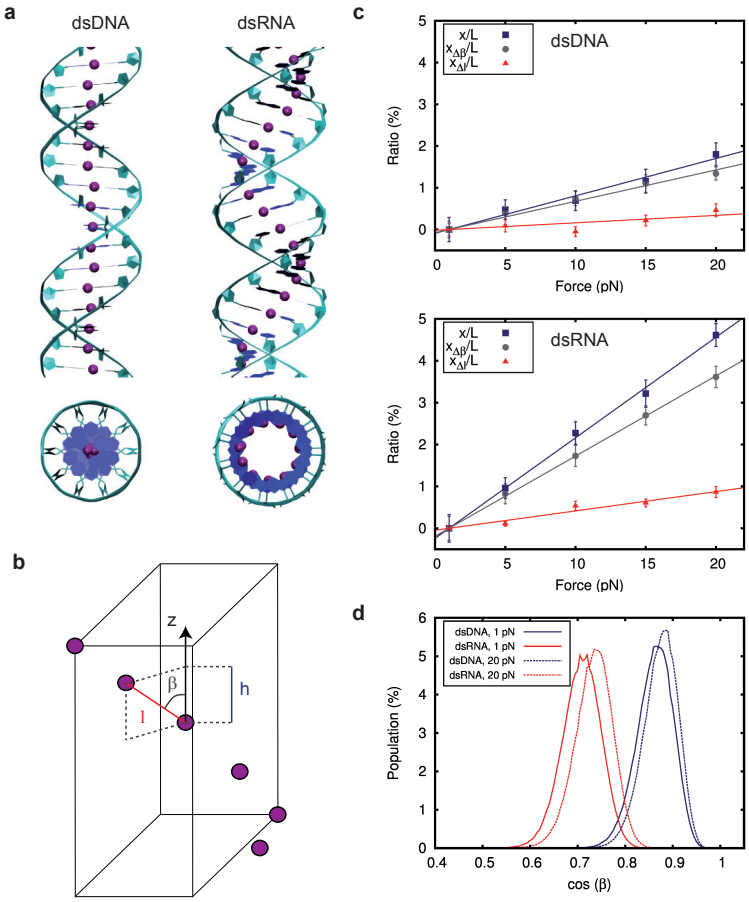

Figure 2.3: A discrete model explains the different stretching response of dsDNA and dsRNA. a, Top and side views of dsDNA (left) and dsRNA (right) molecules, with base pair centers represented by purple beads. b, In a given NA molecule, l is contour length of the chain defined by the purple beads; h is the projection of l on the helical axis; and β is the angle defined by these two. The extension can increase by reducing β (denoted by $x_{\Delta\beta}$) and/or by increasing l ($x_{\Delta l}$). Then, the total elongation can be written as $x \equiv \Delta h = x_{\Delta\beta} + x_{\Delta l}$. c, $x_{\Delta\beta}/L$ and $x_{\Delta l}/L$ contributions to the total relative change in extension x/L (same data as in Fig. 2.1a) for dsDNA (upper panel) and dsRNA (lower panel). A linear fit constrained to pass through the origin point (1,0) was carried out for each data set. From the slopes we calculated $k_{\beta,DNA} = 1330 \pm 50$ pN, $k_{\beta,RNA} = 522 \pm 3$ pN, $k_{l,DNA} = 5600 \pm 1500$ pN, and $k_{l,RNA} = 2170 \pm 140$ pN. d, histogram of $\cos\beta$ values for dsDNA and dsRNA for 1 and 20 pN. Error bars in c were calculated as described in Fig. 2.1.

ones calculated from Fig. 2.1a and Eq.B.6, revealing that our simple *springiness model* captures to a high precision the elastic stretching response of the NAs in this range of forces.

In a system of two springs in series, the softer spring dominates the global elastic response. Our data show that when a NA is stretched under forces below 20 pN, the major contribution to the change in extension is the springy term $x_{\Delta\beta}(F)$. In other words, the lengthening induced from separating the centers of the base pairs is much smaller than the contribution coming from deforming the chain. Although this may be expected for dsRNA based on its more open structure, we show here that this claim also holds for dsDNA, for which the springiness is close to zero (Fig. 2.3d). This contrasts with the intuitive idea that dsDNA stretching should involve base-pair separation due to its nearly straight structure. In fact, a careful inspection of Fig. 2.3c shows that the contribution of $x_{\Delta l}(F)$ to the extension is negligible up to 10 pN. Consequently, the qualitatively different stretching elastic response can be attributed to the $\sim$3-fold larger k_β of dsDNA with respect to dsRNA.

2.4 Mechanism of opposite dsDNA and dsRNA twist-stretch coupling

Experiments and simulations have reported an opposite twist-stretch coupling sign for dsDNA and dsRNA [37, 72], resulting in overwinding of dsDNA and unwinding of dsRNA when the molecules are stretched. If we model the molecules as two strings wrapped around each other, our intuition tell us that they will unwind to extend (Fig. 2.4a). While this is the case for dsRNA ($g > 0$), both experiments and simulations agree that this is not so for dsDNA ($g < 0$). The counter-intuitive nature of this result can be understood if we allow changes in the separation of the two strands as previously suggested [42, 72]. A fixed inter-strand separation imposes the molecule to unwind when stretched (the dsRNA case). If the distance between strands is allowed to shrink, then elongation can proceed by overwinding the molecule (the dsDNA case) (Fig. 2.4a).

A direct experimental measurement of the inter-strand separation is very challenging because it would require knowledge of the microscopic details of

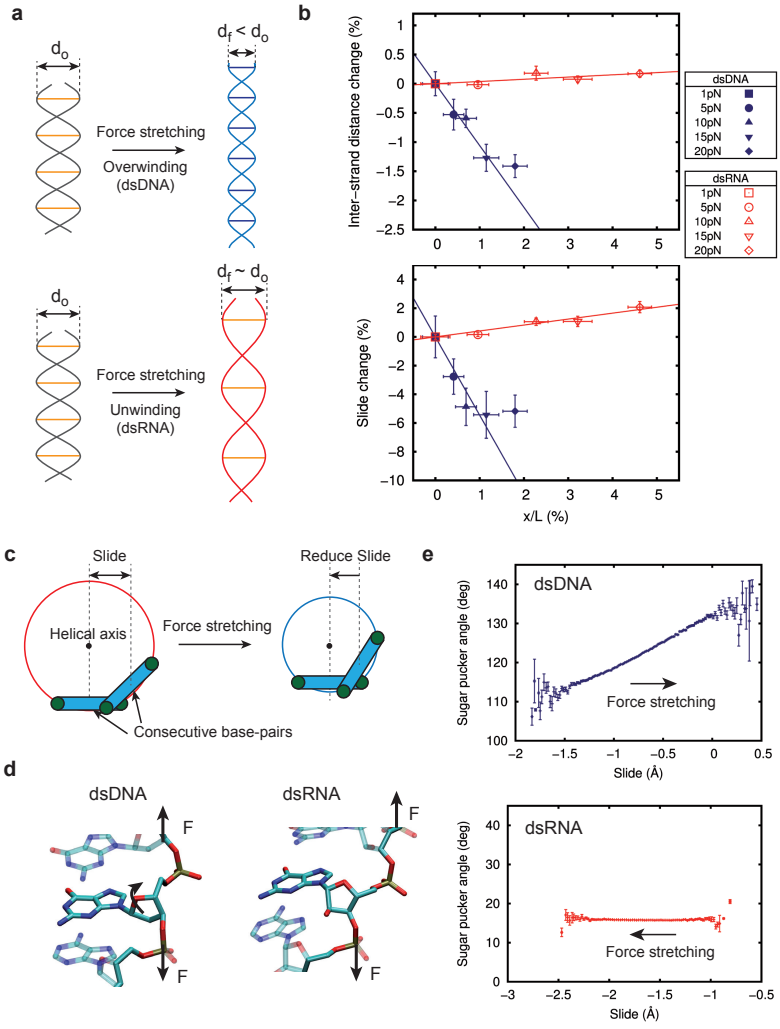

Figure 2.4: Mechanism of the opposite sign of dsDNA and dsRNA twist-stretch coupling. (a, Upper) A double-helix can overwind when stretched if the interstrand distance is able to shrink. (a, Lower) Alternatively, a fixed interstrand distance duplex will unwind when stretched. (b) Relative change of the interstrand distance (Upper) and slide (Lower) plotted against the relative elongation. Datasets were fitted to a linear function, excluding the value at F = 20 pN for dsDNA. Error bars were calculated as in Fig. 2.1. (c) The reduction of slide in dsDNA is linked to a reduction of the interstrand distance (d) The orientation of the sugar with respect to the phosphate backbone is different for dsDNA (Left) and dsRNA (Right). (e) Fluctuations in sugar pucker angle with respect to slide. The bin size is 0.02 Å. Errors are the SEM.

the system. Our constant-force μs-long simulations allowed us to look at the atomistic level and directly measure the force dependence of the inter-strand separation. We measured this parameter as twice the mean distance from the P atoms to the helical axis for each frame and averaged over our simulation time. Interestingly, the evolution of this parameter with the force was very different for both molecules (Fig. 2.4b, upper panel). While the inter-strand separation barely changed for dsRNA when stretched, this value experienced a significant decrease for dsDNA (Fig. 2.4b, upper panel). This supports the idea that overwinding or unwinding is coupled to the change in inter-strand separation as suggested above.

Inspection of the variation of all base pair step parameters with the force, showed a striking qualitative difference between dsDNA and dsRNA in the slide parameter (Fig. 2.4b, lower panel,and Fig. 2.5). The slide parameter represents the displacement of two consecutive base pairs along the y-axis (that defines the direction of the base pairing), and it is negative most of the time for both molecules. For the sake of clarity, we consider a decrease of slide as a decrease of its absolute value, which corresponds to the approaching of consecutive base pairs. Importantly, the inter-strand distance and the slide are strongly correlated, *i.e.*, a decrease of slide leads to a decrease of inter-strand distance (Fig. 2.4c). For dsDNA, the slide decreased when stretched, while for dsRNA it barely changed (Fig. 2.4b, lower panel). Consequently, dsDNA overwinding, *i.e.* inter-strand distance reduction, can be explained in terms of the reduction of its slide parameter as it elongates (Fig. 2.4c). Note, however that both the inter-strand separation and slide of dsDNA deviated from the linear behavior at F =20 pN (see next section). On the contrary, dsRNA was not able to reduce its inter-strand distance/slide and therefore it can only elongate by unwinding.

The different behavior of the slide parameter of dsDNA and dsRNA can be traced down to the most fundamental difference between these two molecules. The connection of the nitrogenous bases to the phosphate backbone is done through a deoxyribose sugar in dsDNA and a ribose sugar in dsRNA. The additional oxygen atom of the ribose sugar results in a distinct orientation with respect to the backbone (Fig. 2.4d) [114]. Moreover, application of force through

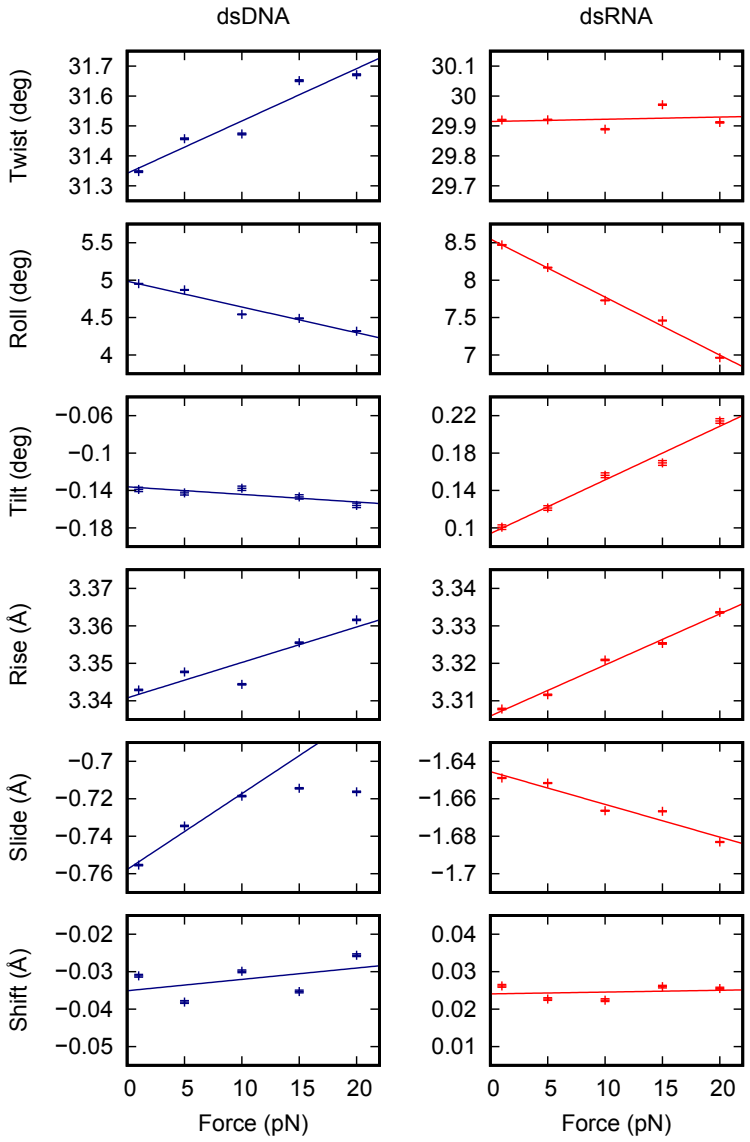

Figure 2.5: Inter base pair parameters as a function of the force for dsDNA (blue) and dsRNA (red). Error bars are the standard error of the mean. To guide the eye, a linear fit was plotted for all the base pair parameters. Concerning the slide of dsDNA, only the first three points were used in the linear fit.

the phosphate backbone is likely to affect the particular stereochemistry of the sugar. We measured the sugar pucker angle in our simulations and obtained an homogeneous population of C3-endo ($\sim$ 18 deg) for dsRNA, and a more disperse population with a average value of $\sim$ 122 deg for dsDNA. From the fluctuations of the molecules we observed that while the sugar pucker angle of dsDNA significantly changed with the slide, it remained constant around the C3-endo configuration in dsRNA (Fig. 2.4e). This larger flexibility of the sugar pucker angle in dsDNA provides an additional degree of freedom to decrease the slide allowing the molecule to overwind when stretched. In the dsRNA case, the lack of this degree of freedom precludes reduction of the inter-strand separation resulting in unwinding upon stretching.

2.5 DsDNA and dsRNA twist stretch coupling dependence with force

Finally, in order to understand the dependence of g with the force, we analyzed the fluctuations of the system. The values of the helical rise at each constant force simulation were discretized in bins of 0.02 Å and the mean helical twist and slide were computed for each bin (Fig. 2.6a, b). This representation revealed that the coupling between twisting and stretching defines a more complex scenario than a simple straight line with regions where the molecule overwinds and others where it unwinds. It also emphasized the tight relation between the slide and the helical twist.

In the case of dsDNA there were two clearly defined regions (Fig. 2.6a). For helical rise $\leq$ 3.3 Å the molecule overwound when stretched (g <0), in accordance with the negative sign previously found for g (Fig. 2.1). For helical rise $\geq$ 3.3 Å (green region), DNA was no longer able to overwind (helical twist remained approximately constant) and the slope in this region was close to zero (gray line in Fig. 2.6, upper panel). As a consequence, molecule excursions into this region lead to a reduction of the average value of g (black line in Fig. 2.6a). As the force increased, the green region became more and more populated (Fig. 2.6c), such that, at F = 20 pN, its population is large enough to induce a decrease in $-g/S$ of 10 $\%$ (Fig. 2.1). The observed flattening in

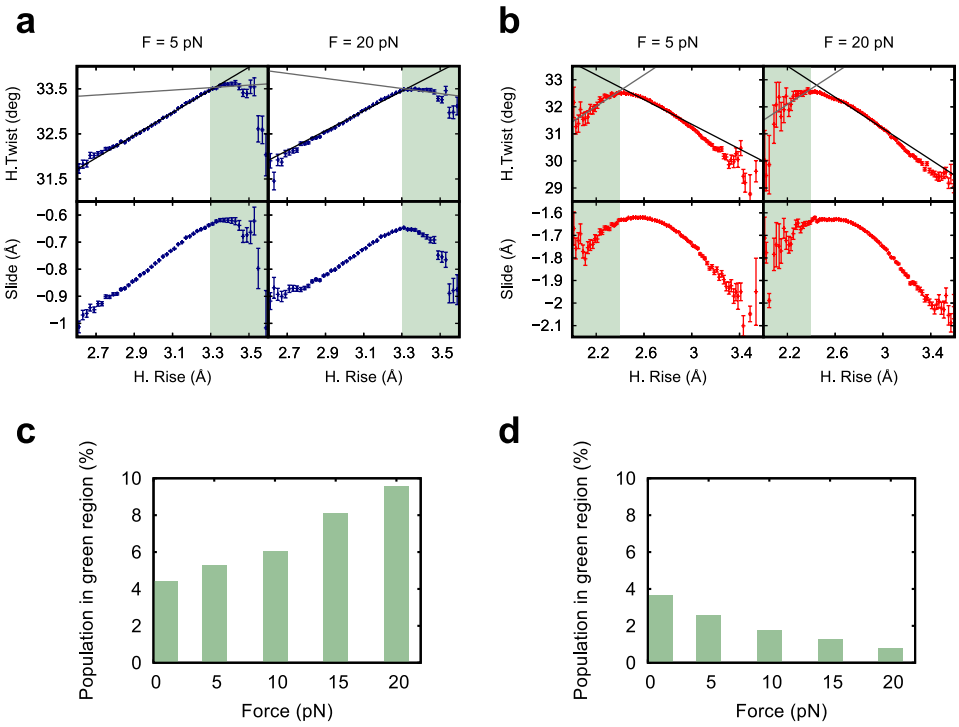

Figure 2.6: Coupling between twist and slide with helical rise as a function of the force. (A) Fluctuations in helical (H.) twist and slide plotted against helical rise at forces F = 5 pN and F = 20 pN for dsDNA. The helical rise was discretized in bins of 0.02 Å, and the mean value of the slide and the H. twist and helical rise were computed in each bin. Error bars are the SEM of each bin. The green region is defined as a helical rise >3.3 Å for dsDNA. The dashed line is the linear fit of the points in the green region, and the continuous line is the fit to all data points. (B) Fluctuations in H. twist and slide for dsRNA. The green region for dsRNA is defined as a helical rise <2.4 Å. (C) Population in the green region for dsDNA at different forces. (D) Population in the green region for dsRNA at different forces.

mean slide at 20 pN (Fig. 2.4b, lower panel) can also now be explained by the contribution of data points in the green region (Fig. 2.6a, lower panels). The same argument applies to the fact that the inter-strand distance barely changed at 20 pN (Fig. 2.4b, upper panel). At higher forces, we predict an inversion in the trend of the slide, inducing an eventual increase in the inter-strand separation and an inversion of the sign of g. Furthermore, additional simulations at 30 pN confirmed the trend of decreasing $-g/S$ observed at 20 pN. These results are in reasonable agreement with reported experimental measurements (Fig. 2.7). This is the first work that provides a possible atomistic explanation for this experimentally observed effect [42, 43, 45, 46].

A similar reasoning can be made for dsRNA (Fig. 2.7c). In this case, when the molecule explored the region with values of helical rise $\leq$ 2.4 Å (green region), it overwound when stretched (g <0, gray line in Fig. 2.6b, upper panels), contrary to the average behavior of dsRNA. Data in the green region contributed to lower the average value of g (black line in Fig. 2.6c). As the force increases, this region gets depopulated (Fig. 2.6d), so that the g value for dsRNA increases (Fig. 2.1c) facilitating the unwinding of the molecule. This consolidates the positive value of g^{RNA}, as opposed to the trend observed for dsDNA. Furthermore, the combined effect of the slope in the green regions together with changes in their population explains why the decrease in $-g/S$ induced by the force observed in Fig. 2.1c could be detected at lower forces for dsRNA ($F \sim$ 5-10 pN) than for dsDNA ($F \sim$ 15-20 pN).

Values of the elastic parameters S and C obtained from our MD simulations were in good agreement with experimental data (Table 2.1). Regarding g, values were in reasonable agreement with recent MD simulations performed at zero force [72], but systematically larger in absolute terms compared to the available experimental reports [37, 42–44] . This is not surprising given that regions of helical rise $\geq$ 3.3 Å for dsDNA and $\geq$ 2.4 Å for dsRNA represent high energy rarely-visited states but, as shown above, they contribute significantly to the value of g. Therefore, it is likely that these higher energy states are not being sampled enough and their contribution to g underestimated, yielding to a higher absolute value for g.

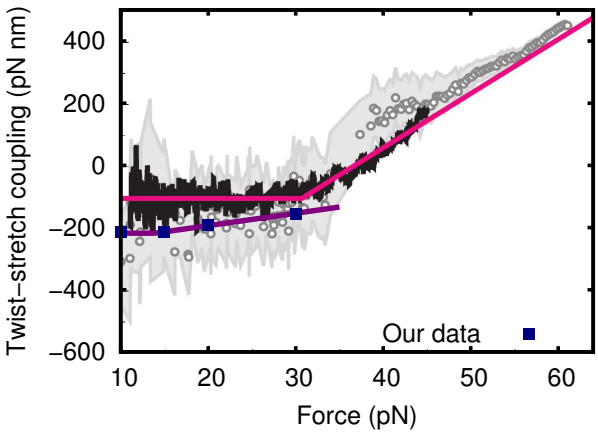

Figure 2.7: Comparison with experimental results of the force dependence of the twist-stretch coupling. Black line: Twist-stretch coupling obtained in [45] from the measurements performed in [42]. Red line: fit of the force-twist curve from [42] using the twistable worm-like chain (tWLC) model proposed in [45]. Grey dots: determination of the twist-stretch coupling from force-extension curve using the tWLC model. Blue squares: values of the twist-stretch coupling measured from our simulations. An additional simulation at $F = 30$ pN was performed to confirm the trend of increasing g. Purple line: fit to our data points following the expression for $g(F)$ proposed in [45], i.e., constant g up to a given force, F_c, and linear for forces above F_c. The linear region was extended up to 35pN. Image adapted from [45], with permission from Macmillan Publishers Ltd: Nature Physics, copyright 2011.

2.6 Conclusions

Our all-atom μs-long MD simulations of dsDNA and dsRNA subjected to stretching forces up to 20 pN allowed us to fully characterize their mechanical response, extracting all the elastic constants, and providing an explanation to the striking differences found in single-molecule experiments. Furthermore, we showed how a hierarchical analysis from a continuum approach, through a discrete model, to an all-atom description paves the way to link the disparate behavior to structural differences in the arrangement of consecutive base pairs, and to the different sliding between base pairs upon stretching. Our work highlights MD simulations as a powerful tool to unveil the connection between forces and structure of nucleic acids and, possibly, gain insight into the associated changes in their biological functionality.

3 Sequence-Dependent Mechanical Response of dsDNA

Adapted from: A. Marin-Gonzalez, J. G. Vilhena, F. Moreno-Herrero, R. Perez, *DNA Crookedness Regulates DNA Mechanical Properties at Short Length Scales*, Phys. Rev. Lett. **122**, 048102 (2019).

3.1 Introduction

The mechanism by which proteins interact with the genome with such extraordinary specificity is still an open question in biology. Since the discovery of the DNA double helix (dsDNA), it became clear the existence of a sequence dependent set of hydrogen bond donors and acceptors that are exposed in the major groove and specifically recognized by certain amino acids. However, there is increasing evidence that this mechanism is far from sufficient. In a number of DNA-protein complexes, DNA adopts a conformation that substantially deviates from the canonical B-form [47, 115, 116], pointing to an indirect readout mechanism by which proteins use the sequence-dependent flexibility of the DNA [115, 117]. Among the most-studied cases are sequence-dependent DNA deformations (such as A-like structures, kinked base pair steps and A-tracts) that play an important role in transcription regulation [118–121].

However, many aspects of DNA flexibility have so far remained elusive. For example, it is not fully understood how a relatively stiff molecule, with a persistence length of $P \sim 50$ nm, is able to wrap around a histone octamer of ~ 4 nm of radius. Even more intriguing is the fact that some sequences are hardly able to form stable nucleosomes, arguably as a consequence of a distinct confor-

mation or mechanical properties [122, 123]. The same question holds for other DNA-protein complexes, in particular, for many repressor systems where a highly bent loop is predicted in the DNA [124]. These observations raise the need of addressing the mechanical properties of DNA from a molecular perspective that accounts for sequence effects.

3.2 The crookedness affects the extension of DNA sequences

Using constant-force molecular dynamics (MD) simulations [125] we observed that the extension of the DNA changed from one sequence to another for molecules with the same number of base pairs. We performed over 1 μs-long MD simulations of 18 base pair long DNA molecules with benchmark sequences of the form CGCG$(NN)_5$CGCG, where NN denotes AA, AC, AG, AT, CG and GG, see Table 3.1. We then computed the average structure of these sequences at 1 pN force and found that the variability in the extension reflects an intrinsic curvature of the molecule. This curvature is apparent if we represent the centers of the base pairs (color beads in Fig. 3.1). Notice that this curvature is the springiness deformation proposed in [37, 48] and developed into the springiness model in Section 2.3. In this Chapter, we will refer to this curvature using the alternative term *crookedness*, in analogy with a crooked road, whose trajectory is not straight. The reason for this choice is to emphasize that, contrary to the springiness, this deformation 1) accounts for sequence effects on DNA flexibility; and 2) establishes a one-to-one correspondence between the DNA structure and flexibility (see below).

Importantly, DNA crookedness is of the order of a few nanometers ($\sim$ 2 nm), a scale comparable with the histone octamer radius [129], the DNA curvature of the DNA-endonuclease I-PpoI complex [130] and several examples of sharply bent DNA found in regulatory regions [124]. This is illustrated in Fig. 3.2. Additionally, it is useful to compare the crookedness with the curvature predicted by the WLC model. Notice that the WLC curvature is entropic (*i.e.* temperature dependent and absent as $T \to 0$), whereas the crookedness is enthalpic and, thus, characteristic of the structure of the molecule and present at zero temperature. Figure 3.2 illustrates that, at length scales of the order of a few nm, the

Benchmark Sequences			
Label	Name	Sequence	Ref
AA	poly-A	AAAAAAAAAA	
CG	poly-CG or CG Island	CGCGCGCGCG	
AG		AGAGAGAGAG	
AC		ACACACACAC	
AT		ATATATATAT	
GG	poly-G	GGGGGGGGGG	
Testing Sequences			
DDD	Drew-Dickerson Dodecamer	CGCGAATTCGCG	[126]
TATA	TATA-element	TATAAAAG	[127]
TFBS	Trans. Factor Binding Site	GGATGGGAG	[120]
G_4CG_4		GGGGCGGGG	
G_4AAG_4		GGGGAAGGGG	
A-tracts			
DUE	DNA Unwinding Element	GATCTATTTATTT	[128]
A_4TA_4		AAAATAAAA	
A_4GGA_4		AAAAGGAAAA	
Test A-tracts			
A_8T		AAAAAAAAT	
A_8GG		AAAAAAAAGG	
mCGmCG			
mCGmCG	Hypermethylated CG Island	mCGmCGmCGmCGmCG	
DNA with all step kinds			
DNA steps	DNA with all step kinds	GCGCAATGGAGTACGC	[72, 125]
RNA with all step kinds			
RNA steps	RNA with all step kinds	GCGCAAUGGAGUACGC	[72, 125]

Table 3.1: Sequences studied in this work. All newly simulated sequences (that is all the molecules except DNA and RNA with all steps [125]) were sandwiched betweeen CGCG handles. mC stands for deoxy-5-methyl cytosine.

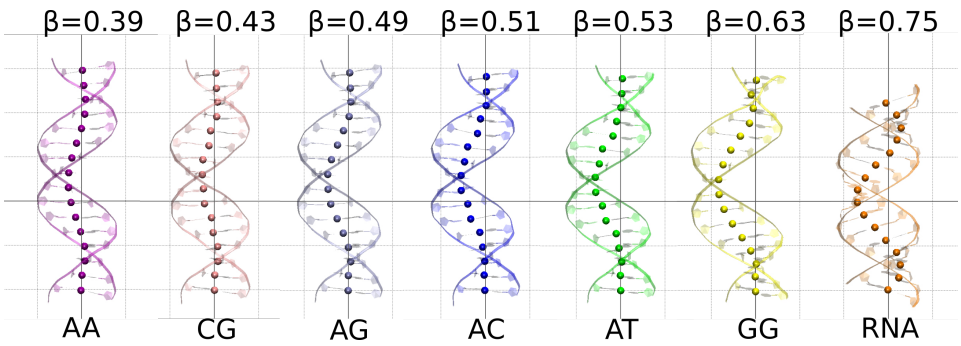

Figure 3.1: Sequence-dependent DNA crookedness. Average structures and computed β values (in rads) at 1 pN of the sequences CGCG(NN)$_5$CGCG with NN=AA, CG, AG, AC, AT and GG. The dsRNA molecule from Chapter 2 was included. The beads represent the centres of the base pairs. The terminal base pairs have been omitted. Grid size is 1 nm.

curvature predicted by the WLC ($\sim$ 16 nm of radius) is much smaller than the crookedness curvature intrinsic to the DNA molecule. This contrasts with the behaviour expected for DNA molecules longer than the persistence length, for which entropic effects are no longer negligible and will eventually dominate over the crookedness.

Crookedness will be quantified in a similar way to the springiness (Eq. 2.3), by expressing the contour length of the center-line polymer as the sum of distances between consecutive base pair centers. That is,

$$l = \sum_{i}^{N-1} l_i, \tag{3.1}$$

where N is the number of base pairs of the molecule and the sum runs over all the base pair steps. The crookedness parameter then reads

$$\cos\beta \equiv \frac{h}{\sum l_i}, \tag{3.2}$$

being h the extension or end-to-end distance. When the molecule is completely straight, the line that runs through the base pairs is perfectly aligned, $h = \sum l_i$, and the crookedness, β, is zero. As this line deviates more from the helical axis, the ratio $h/\sum l_i$ becomes smaller and therefore DNA crookedness increases.

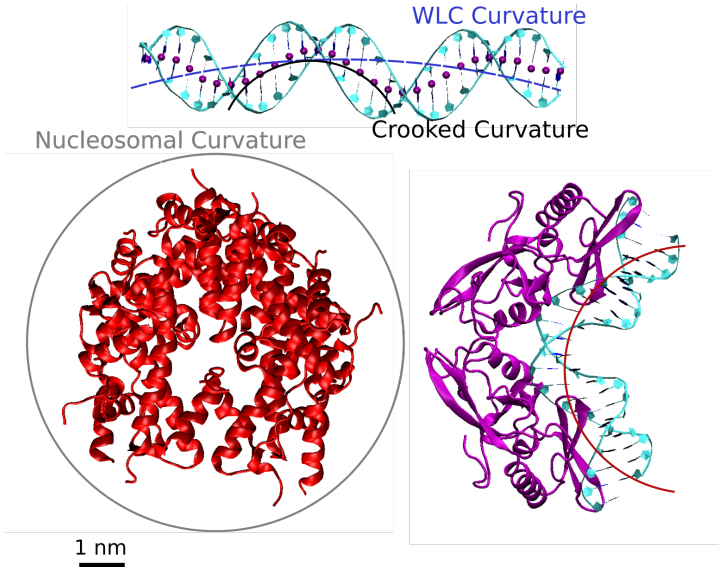

Figure 3.2: Comparison of DNA crookedness with other curvatures. Top: average structure over 250 ns MD of a 30 bps poly-G DNA molecule. The solid black line represents the crookedness curvature and the dashed blue line an estimation of the curvature predicted by the WLC. Bottom: examples of highly curved DNA when bound to proteins. (left), histone octamer crystallized in [129] (PDB ID: 1AOI), where the histone tails have been removed for clarity. A grey circle of radius 41.8 Å represents the trajectory of nucleosomal DNA. (right), crystal structure of the homing endonuclease I-PpoI DNA complex taken from [130] (PDB ID: 1A73), with an estimation of the DNA curvature represented by the solid red line.

B-DNA molecules exhibit β values centered on 0.5 (Fig. 3.1). Moreover, the β for a poly-G sequence of 30 bps (β=0.620 rad) was very similar to the one found for the 18 bps molecule (β=0.626 rad) suggesting that β values are likely to persist for longer DNA molecules.

3.3 Rationalizing the sequence-dependent DNA stretching flexibility

Additionally to sequence-dependent conformations, proteins often exploit DNA flexibility [115, 117]. Therefore, a complete comprehension of the biological relevance of DNA crookedness requires understanding its effect on DNA mechanical properties. We propose a model to rationalize the relation between β and DNA stretch modulus, S [1]. This model is a generalization of the springiness model from Chapter 2. A detailed derivation of the model is presented in Appendix C.

3.3.1 Basic features of the model

From the definition of the crookedness (Eq. 3.2), the extension of a DNA molecule can be written as

$$h(F) = \cos\beta(F) \times \sum_{i=1}^{N-1} l_i(F). \tag{3.3}$$

Analogously to Section 2.3, we can decompose the molecule elongation, $x(F)$, as a sum of $N-1$ *local* contributions coming from elongating individual base pair steps, $x_{\Delta l,i}(F)$; and the *global* contribution of aligning the base pairs with the helical axis, $x_{\Delta\beta}(F)$. This idea is represented schematically in Fig. 3.3. Mathematically,

$$x(F) = \sum_{i=1}^{N-1} x_{\Delta l,i}(F) + x_{\Delta\beta}(F). \tag{3.4}$$

[1] In the remaining part of this Thesis we will make no distinction between the stretch modulus and the effective stretch modulus (contrary to Chapter 2). We will refer to this mechanical parameter as stretching stiffness, stretch modulus or stretching flexibility indistinctively, and we will denote it as S.

Assuming that these deformations are linear with the force (see Appendix C) we can define their associated elastic constants as

$$k_{l,i} \equiv \frac{LF}{x_{\Delta l,i}(F)} \qquad\qquad k_{\beta} \equiv \frac{LF}{x_{\Delta \beta}(F)} \tag{3.5}$$

where $L \equiv h(0)$. And the stretch modulus is equal to

$$S^{-1} = \sum_{i=1}^{N-1} k_{l,i}^{-1} + k_{\beta}^{-1}. \tag{3.6}$$

According to this equation, the stretching stiffness of a DNA molecule is determined by N parameters, $k_{l,i}$ $(i = 1, \ldots, N-1)$ and k_{β}. In what follows we will show that these parameters can be univocally determined from the DNA sequence and structure. In other words, Eq. 3.6 allows us to determine the stretch modulus of any given DNA sequence by solely *looking* at its equilibrium conformation. Figure 3.3 is a schematic representation of Eq.3.6, showing that the stretching response of a DNA molecule is being modelled as a set of N springs in series with elastic constants $k_{l,i}$ and k_{β}.

3.3.2 Quantifying the local contribution to DNA stretching

Starting from Eq. 3.5, and using the definition of $x_{\Delta l_i}$ (Eq. C.12) and β (Eq. 3.2) one arrives at

$$k_{l,i} = \left(\sum_{j=1}^{N-1} l_j(0) \right) \times \tilde{k}_{l,i}, \tag{3.7}$$

with

$$\tilde{k}_{l,i} \equiv \frac{F}{\Delta l_i(F)} \tag{3.8}$$

This equation illustrates that $k_{l,i}$ is nothing but the stiffness of separating two consecutive base pairs, $\tilde{k}_{l,i}$, multiplied by a prefactor. We may argue that this $\tilde{k}_{l,i}$ is sequence dependent, since it is closely related with the base pair stacking interactions. Moreover, for computing $\tilde{k}_{l,i}$ we will resort to the nearest neighbour approximation, that is, we will assume that this parameter is solely dependent on the base pairs composing the step. Accordingly, there will be ten different values of $\tilde{k}_{l,i}$, corresponding to the ten different dinucleotides or *step kinds*.

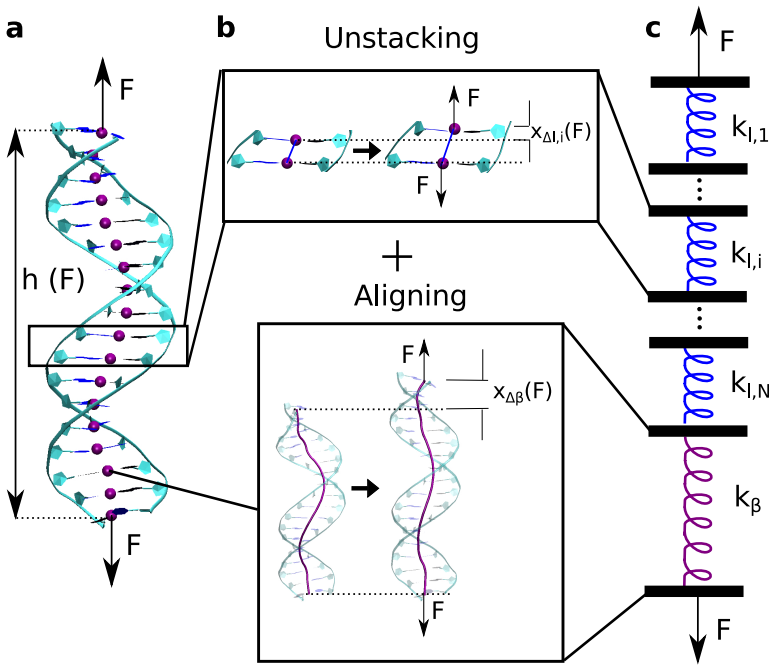

Figure 3.3: A model to link DNA crookedness with DNA stretch modulus, S. (a) DNA molecule with the base pair centers represented by purple beads. An external force, F, induces a change in DNA extension, $x(F)$. (b) DNA can elongate by separating consecutive base pairs (top; $\Delta x_{l,i}$) or by aligning the base pair centers with the helical axis (bottom; Δx_β). (c) DNA is modelled as a set of N springs in series. The first N-1 springs account for the stiffness of elongating individual base pair steps, $k_{l,i}$, and the N^{th} spring is the crookedness stiffness, k_β.

Step	AA	AC	CA	AG	GA	AT	TA	CG	GC	GG
$\tilde{k}_{l_i}$	1820	1660	1370	2390	1230	5500	1990	1340	2350	3500

Table 3.2: The values of $\tilde{k}_{l,i}$ of all the dinucleotides are given in pN/Å. These values were computed as the inverse of the slopes of the linear fits of Fig 3.4.

In order to determine the sequence dependence of these $\tilde{k}_{l,i}$, we ran four additional constant-force MD simulations at F = 5, 10, 15 and 20 pN using the benchmark sequences described above. Note that each of the ten *step kinds* is present at least four times in this set of sequences. We first computed the mean base pair step separation for each step kind at different forces, $l_i(F)$. This separation was obtained using the 3DNA software as $l_i = \sqrt{Slide^2 + Shift^2 + Rise^2}$ [27]. Then we plotted $\Delta l_i(F)$ as a function of the force taking the F = 1 pN simulation as reference, see Fig 3.4. These data sets showed a linear dependence with the force, supporting the assumption made in Eq. 3.5 that this deformation is elastic in this range of forces [2]. Following Eq. 3.8, the values of $\tilde{k}_{l,i}$ are obtained as the inverse of the slopes of the linear fits to these datasets. Knowing the $\tilde{k}_{l,i}$ for all step kinds, one only needs to measure the sum of base pair distances at zero force, $\sum_j^{N-1} l_j(0)$, to obtain $k_{l,i}$ (see Eq. 3.5).

In Table 3.2 we show the computed values of $\tilde{k}_{l,i}$. As anticipated, this parameter is highly dependent on the step kind. In particular, a closer inspection at Table 3.2 reveals that the AT step is the stiffest, in line with previous works which coincide in that this step is the least flexible [47, 66, 131]. In contrast, the GA, CG and CA steps showed the smallest $\tilde{k}_{l,i}$ values. This result is in agreement with a study on DNA crystal structures [47], where these three steps showed the highest standard deviation of the *rise* parameter. Moreover, we reproduced the tendency of *pyrimidine-purine* steps being generally the most flexible, fol-

[2] Only the GG step showed a larger dispersion from the linear response. This is probably due to convergence issues, since this step has an unusually high value of the slide parameter, which is highly variable and strongly affects the value of l_i. We extended our simulations to 2 μs and ran additional simulations at 12 and 18 pN and this deviation from linearity persisted. It is unlikely that we will achieve convergence for this step in the μs timescale.

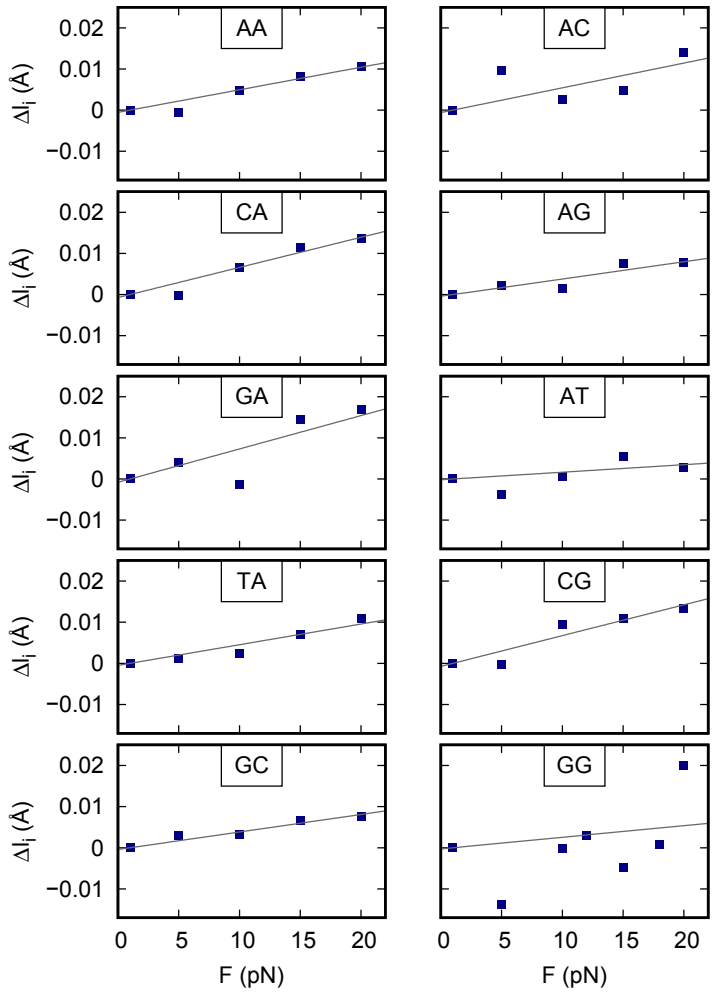

Figure 3.4: Force-induced base pair step elongation of the ten dinucleotide step kinds. The base pair step separation, $l_i(F)$ was computed for each step and averaged over the steps of the same kind and over the 1 μs simulation time for our benchmark sequences at each constant force simulation. We represented the elongation $\Delta l_i(F)$ with respect to the $l_i(0)$ value, taken at 1 pN force, as a function of the applied force. The data sets were fitted to a linear function constrained to go through the (1,0) point. The inverse of the slopes are the $\tilde{k}_{l,i}$ of each step kind given in Table 3.2.

lowed by *purine-purine* and *purine-pyrimidine* [131, 132]. Interestingly, this same trend was recently reported for the dinucleotides stacking energies, where *purine-pyrimidine* interactions are mostly among the strongest and *pyrimidine-purine* among the weakest [133].

3.3.3 Quantifying the global contribution to DNA stretching

Having determined the contribution of the local base pair step elongation ($k_{l,i}$) to the stretch modulus, it remains to measure the global deformation associated with the crookedness k_β (see Eq. 3.6). We will measure this quantity in the same way as the springiness deformation in Section 2.3. Therefore, similar to Fig. 2.3, we measured the sum of local contributions ($x_{\Delta l} = \sum_{i=1}^{N-1} x_{\Delta l,i}$, see Eq. 3.4 and Eq. C.6) and the crookedness contribution ($x_{\Delta\beta}$) to the elongation (x) of the benchmark sequences. This is represented in Fig. 3.5.

These measurements illustrate two important features. Firstly, the DNA stretch modulus is highly dependent on the sequence, as evidenced from the variability in the slopes of the $x(F)/L$ data. Secondly, in line with the findings from Section 2.3, the crookedness deformation, $x_{\Delta\beta}$, was the dominant contribution to the elongation in all the studied sequences. Hence, in order to quantitatively predict the DNA stretching flexibility, we need to accurately capture the sequence dependence of k_β. Given that this parameter represents a global deformation of the molecule (see Fig. 3.3), a local dinucleotide approach similar to $k_{l,i}$ is likely to be insufficient.

We propose that this global deformation depends on the *overall* structure of the molecule, concretely on the crookedness itself. Figure 3.6a shows the calculated values (blue points) and the extraordinary fit provided by the function

$$k_\beta(\beta) = Ae^{-k\beta} + B \tag{3.9}$$

with parameters A = (2.24 ± 1.24) × 10^6 pN, B = 700 ± 120 pN and k = 16.2 ± 1.5. Remarkably, this phenomenological function reproduces the results for 11 additional sequences exhibiting a broad range of β values (Table 3.1) [3]. This reveals an exceptional property of DNA crookedness: the relation with

[3] Our choice for the fitting function is purely phenomenological. Other functions, such as

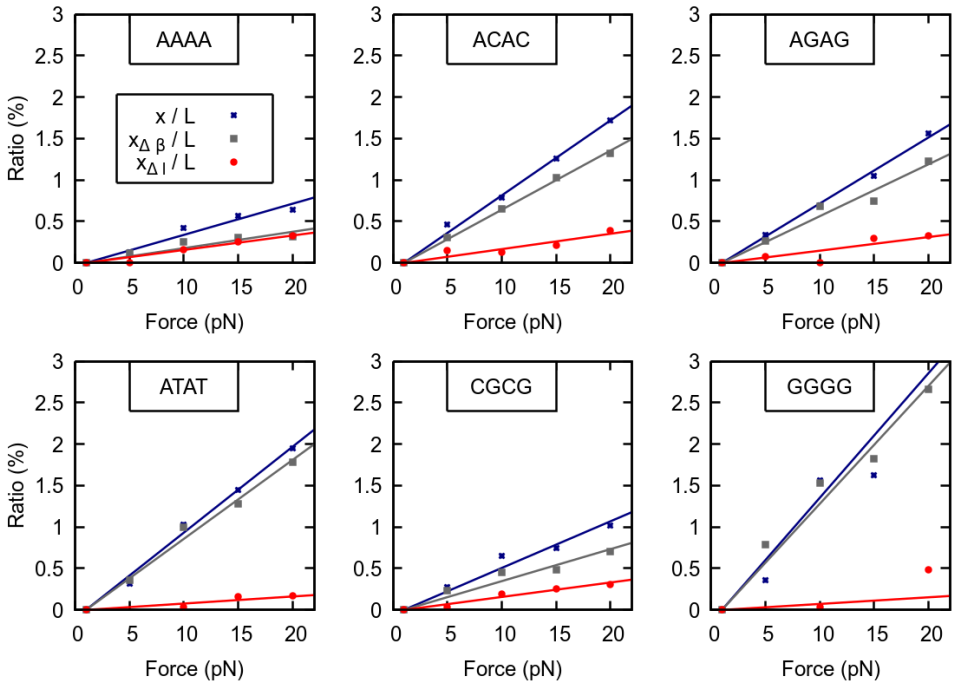

Figure 3.5: Contributions to the force-induced elongation of the benchmark sequences. The total elongation, Δx; the contribution to the elongation coming from aligning the base pairs with the helical axis, Δx_β; and the contribution coming from elongating base pair steps, $\Delta x_l \equiv (\sum \Delta x_{l,i})/N$ were computed from the MD simulations of our benchmark sequences. The ratio of these quantities and the molecule extension at 1 pN force, x_0 are represented as a function of the applied force. All the quantities were computed using the *3DNA software* in the same way as described in [125]. The inverse of the linear fits of $\Delta x(F)/x_0$ and $\Delta x_\beta(F)/x_0$ yield respectively the stretch modulus, S, and the crookedness flexibility, k_β, that are shown in Fig. 3.6.

its associated stiffness is bijective. In other words, an *equilibrium* structural parameter, β, univocally determines a *dynamical* response, k_β; and vice versa.

The one-to-one correspondence between β and k_β provides predictive power to our model. Indeed, if the equilibrium structure of a DNA sequence is known (by NMR, crystallography, MD, ...) one can measure β and the base pair separation distances. Then, k_β can be obtained from Eq. 3.9 and $k_{l,i}$ can be computed from Table 3.2 and $\sum l_i$ according to Eq. 3.7. From the values of k_β and $k_{l,i}$, one can compute S using Eq. 3.6. Figure 3.6b confirms the good agreement with our predictions and our measured values for a wide range of S values ($\sim$ 800-3000 pN). Importantly, our measurements are in line with previous single-molecule experiments (see Fig. 3.6b) and MD simulations [56, 135].

3.4 Discussion

In addition to predicting the stretch modulus, our model provides valuable information about the nature of its sequence dependence. Our results show that S is not dependent on the GC content, but rather, on how guanines and cytosines are distributed along the strands. Molecules with alternating GC's are much less prone to stretching than those where several guanines are placed sequentially on the same strand. This adds to other examples where mechanical stability is independent of thermal stability [136].

Notice the extremely variable stretching flexibility of molecules that are rich in poly-A regions, such as A_8T and A_4TA_4. Sequences consisting of four or more consecutive A's and T's are called A-tracts and are known to introduce anomalies in DNA mechanical properties [62, 89, 119, 137]. When inserted in an arbitrary sequence, short A-tracts induce a curvature in the DNA. We claim that this is the case in A_4TA_4 and A_4GGA_4, where the A-tract induced curvature might be responsible for the enhanced stretching flexibility in these molecules (Fig. 3.6a void green triangles). On the contrary, long A-tracts appear to be rigid and have been associated with nucleosome depletion *in-vivo* [122]. Consistent with this, we found that A_8T and A_8GG, with a 8 bp-long A-tract show a very large

$A\beta^{-n} + B$ with $n = 6, 7$, also fit to the data. We chose the function $Ae^{-k\beta} + B$ because it is smooth and reasonably easy to converge in conventional fitting programs.

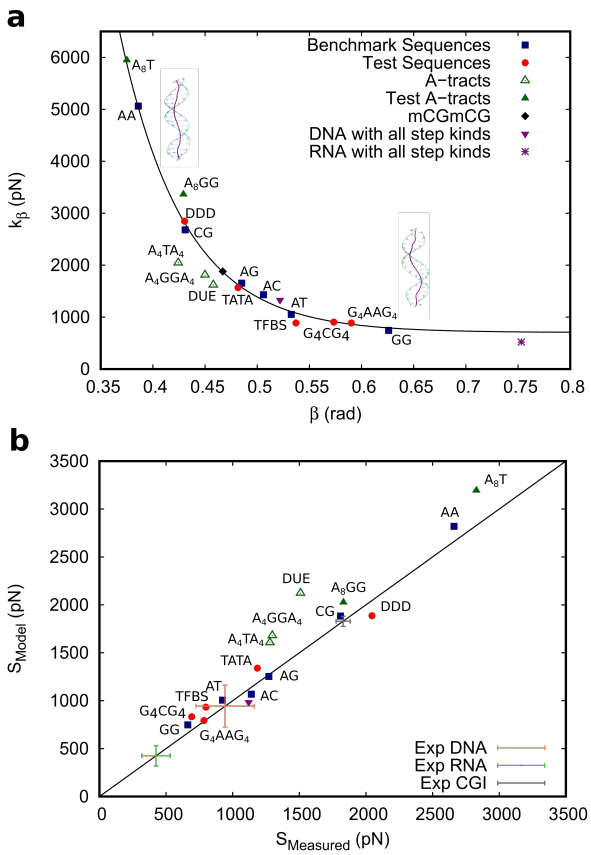

Figure 3.6: A predictive model of DNA stretch modulus. (a) k_β values as a function of the crookedness, β. k_β was computed using Eq. C.7 and Eq. C.5, taking the F = 1 pN simulation as reference. β was computed from the F = 1 pN simulation using Eq. 3.2. The blue squares are the points used for the fit $k_\beta(\beta) = Ae^{-k\beta} + B$. The rest of the sequences provide a test to the model (see Table 3.1 for a list of the simulated sequences). (b) The computed value of S is plotted as a function of the value directly measured from the force-extension curves. The stretch modulus was computed from our model for the simulated sequences described above using Eq. 3.6, the value of the fit of $k_\beta(\beta)$, Table 3.2 and the values of β and $\sum l_i$ obtained from the F = 1 pN simulation. The experimental value of the S for arbitrary DNA and dsRNA sequences are the average value of the measurements from references [32, 34, 36, 37, 111] and [32, 37] respectively; and the error bars are the standard deviation of these values. The experimental value of the CpG Island is the one from [134] with the error reported there.

stretching rigidity (Fig. 3.6a solid green triangles). The paradoxical mechanical properties of A-tracts will be discussed in detail in Chapter 5.

Another interesting sequence is that consisting of alternating C's and G's. About 70% of annotated gene promoters are associated with so-called CpG islands (CGI), rich in CpG steps [123]. Notice that our CG-alternating sequence shows the largest proportion of CpG steps possible in any molecule: one in every two steps. As in the case of the poly-A, CGI have been attributed to nucleosome destabilization [138]. These findings together with Fig. 3.6a suggest a possible relation between an unusually high crookedness stiffness and nucleosome destabilization. Hypermethylation of CGI commonly induces gene silencing and in some cases this has been attributed to nucleosome stabilization [123]. We found that complete hypermethylation of the poly-CG molecule significantly increases its crookedness and stretching flexibility, in quantitative agreement with recent optical tweezers experiments reporting both an unusually high S and a softening induced by hypermethylation in CGI [134]. This could increase nucleosome affinity for a hypermethylated CGI [139]. Notice however, that outside this context, methylation is known to reduce DNA flexibility and destabilize nucleosomes [140–142].

We have shown that DNA crookedness can be tuned by specific sequences, being the main responsible for a sequence dependent stretching flexibility. In addition, the crookedness mechanism, which regulates DNA enthalpic bending at short scales (10 bp), might extend to longer length scales. This would be similar to the case of A-tract induced bending. Although a single A-tract shows a small kink at the 10 bp scale, placing several A-tracts in phase with the helical pitch results in highly bent DNA molecules longer than 100 bp ([89] and Chapter 5). Similarly, one could think that placing alternating high and low crookedness motifs, such as poly-G's and poly-A's, would result in highly bent molecules. Supporting this idea, G-rich and A-rich motifs, with a periodic repeat of 10 bp and in anti-phase with each other, are frequently found in nucleosome positioning sequences [143–145]. This crookedness regulation of DNA curvature is illustrated in Figure 3.7a. This mechanism would add to other DNA features such as nicks [146, 147], mismatches [147] or kinks [148, 149] that are known

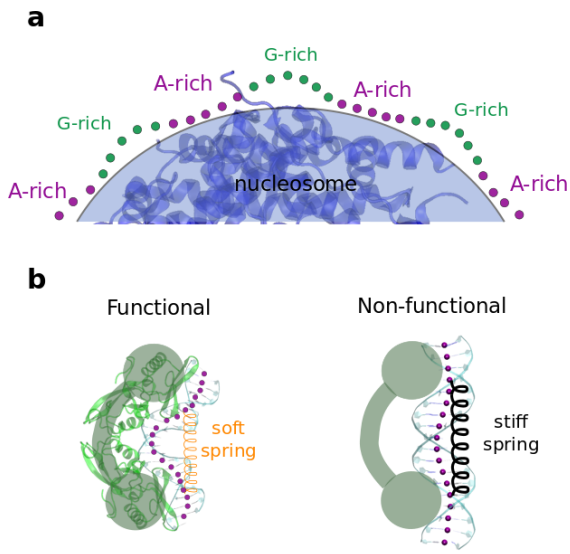

Figure 3.7: Implications of DNA crookedness on DNA-protein interactions. (a) The base pair centres chain of a DNA molecule are represented by color beads. We propose that net directional bending could be achieved by alternating high and low crookedness motifs, such as G-rich and A-rich sequences. (b) Additionally, crookedness flexibility could be exploited to induce a conformational change in DNA towards an A-like structure. On the left is the crystal structure of the I-PpoI DNA complex [130] (PDB ID: 1A73) , the same as in Fig. 3.2. When bound to this protein a high distortion is found on the DNA, which would be more favourable for highly crooked and flexible sequences (left) than for sequences for which this flexibility is hindered (right).

to facilitate DNA bending at short scales.

Additionally, highly crooked molecules might be preferred for DNA-ligand binding where DNA flexibility is required. Indeed, interaction of DNA with proteins and drugs commonly modifies the B-DNA to a more A-like form and this occurs in a sequence-dependent manner [116]. According to Fig. 3.6, molecules with high β would be more prone to form an A-type helix when bound to a protein both because their structure is already closer to the A-form and because of their enhanced flexibility (Fig. 3.7b). Finally, crookedness may modulate the charge distribution along the duplex, known to be a mechanism for protein-DNA recognition by electrostatic interactions [121].

3.5 Conclusions

In this work we have introduced the crookedness: a structural code imprinted in the DNA sequence that modulates its local curvature. Our extensive all-atom simulations allowed us to unveil a striking one-to-one correspondence between this structural parameter and its associated flexibility. Such remarkable correspondence was exploited to build a discrete model to compute the stretch modulus of any DNA sequence given solely its equilibrium structure. Our results contribute to rationalize the idea that DNA sequence regulates the local curvature and the mechanical properties of the double helix at the length scale relevant for biological function.

4 Sequence-Dependent Mechanical Response of dsRNA

Adapted from: A. Marin-Gonzalez, J. G. Vilhena, F. Moreno-Herrero, R. Perez, *Sequence-dependent mechanical properties of double-stranded RNA*, Nanoscale **11**, 21417 (2019).

4.1 Introduction

Double-stranded RNA (dsRNA) performs a wide variety of functions inside the cell. For example, dsRNA molecules carry the genetic information in some viruses, trigger gene silencing or activate the immune response against viral attacks [11, 12, 150–152]. In addition, dsRNA helices are ubiquitous in the 3D structure of regulatory RNAs and in the ribosome [14, 15, 153, 154]. Indeed, the formation of double-helices is often a prerequisite for the folding of RNA into complex tertiary and quaternary structures [155, 156]. Not surprisingly, it has been estimated over one half of the nucleotides in structured RNAs are engaged in canonical Watson-Crick (WC) base pairing [157, 158].

Many of the biological processes involving canonical dsRNA helices interrogate the mechanical properties of the duplex. This occurs both at a global scale – *e.g.* during packaging of a kilo-base-pair long dsRNA molecule inside the viral capsid – and at a local level – due to proteins that distort the dsRNA structure over distances of a few base pairs [159–161]. Furthermore, the formation of tertiary RNA structures involving contacts between canonical dsRNA helices are greatly affected by the sequence-dependent flexibility of these duplexes [158]. Therefore, a complete understanding of sequence-dependent dsRNA flexibility

might pave the way to designing complex 3D structures from canonical helices with well-characterized mechanical properties.

Recent single molecule experiments have assessed the mechanical properties of long dsRNA molecules of random sequence, revealing two striking differences with DNA [32, 37, 99]. Firstly, dsRNA stretches $\sim$ 3 times more under an external force than its dsDNA counterpart [32]. Namely, dsRNA has an effective stretch modulus of $\sim$ 400 pN, much lower than that of dsDNA ($\sim$ 1200 pN). Secondly, dsRNA unwinds upon elongating [37], whereas dsDNA overwinds when stretched [42, 43]. Nevertheless, an important aspect of dsRNA flexibility remains elusive: how its mechanical response depends on the nucleotide sequence.

Molecular dynamics (MD) and Monte Carlo (MC) simulations are an excellent complement to single-molecule methods in the study of the mechanical properties of nucleic acids (NA) and allow in depth exploration of sequence effects [48, 52, 56, 58, 66, 74, 75, 162]. In the case of dsDNA, extensive MD studies have unveiled and characterized a complex scenario of sequence-dependent mechanical properties [52, 56, 66, 162]. However, in contrast to the vast literature on dsDNA flexibility, much less is known about how the sequence affects the mechanical properties of dsRNA. Based on 150 ns-long MD simulations, Faustino et al concluded that sequence patterns of dinucleotide flexibility are reasonably similar for dsRNA and dsDNA [58]. Nevertheless, this local approach lacked a systematic analysis of sequence effects on the overall mechanical response of the duplexes. In parallel, a recent MC simulation based on crystallography-derived structural parameters explored sequence effects on the global flexibility of long RNA and DNA duplexes [48]. Although this work provided insightful predictions of the mechanical properties of these molecules, it was unable to reproduce the characteristic opposite twist-stretch coupling of dsDNA and dsRNA.

In this work we study the sequence-dependent mechanical properties of dsRNA by using constant-force molecular dynamics (CFMD) simulations. This approach was previously shown to reproduce the experimental mechanical parameters of dsRNA (see Chapter 2). Interestingly, when applied to dsRNA molecules consisting of repeating dinucleotides, our microsecond-long CFMD simulations

revealed a strongly sequence-dependent mechanical response. A thorough analysis allowed us to identify the high local flexibility of pyrimidine-purine steps as a critical factor in modulating the global elastic response of these duplexes. Consistently, when inserted in dsRNA molecules of random sequences, these pyrimidine-purine steps softened the mechanical response of the entire helices. These results motivate the exploration of sequence effects on dsRNA flexibility by means of single-molecule manipulation.

4.2 DsRNA mechanical properties are sequence-dependent

The sequence-dependent mechanical properties of dsRNA were studied using CFMD simulations (see Section 1.5.1). We first considered six duplexes with sequences $G_4(NN)_8\ G_4$, where NN=AA, AC, AG, AU, CG, GG (see Table 4.1). Five CFMD simulations were run for each of the sequences at forces F=1, 5, 10, 15, 20 pN. Further details on simulations and data analysis are provided in Section 1.5.1.

As a measure of the mechanical response of the RNA duplexes, we obtained the force-induced change in extension and helical twist divided by the extension at $F = 1$ pN, as done in Fig. 2.1. These data are represented in Fig. 4.1 and reveal three important features. Firstly, for all studied sequences, both the elongation and change in helical twist show a linear dependence on the force, in agreement with the elastic rod model (Eq. 2.2, 2.3). Secondly, as evidenced by the negative slopes of Fig. 4.1b, all sequences show the unwinding-when-stretched behavior characteristic of dsRNA [37] (see Chapter 2). Finally, we found that the mechanical response of dsRNA is strongly affected by the sequence.

Interestingly, a very similar trend is observed in both the extension and twisting response of the RNA duplexes under force. Namely, sequences that are more stretched under an external force (larger slope in Fig.4.1a) are more prone to force-induced unwinding (Fig. 4.1b). This finding is consistent with previous works reporting simple deformability patterns in dsRNA [61, 163]. According to their force response, the poly-CG is the softest molecule, followed by the poly-AC (see Fig. 4.1). On the other hand, the poly-A, poly-G and poly-AG are the

Benchmark Sequences			
Label	Sequence	% GC	# YR Steps
Poly-CG	$G_4(CG)_8G_4$	100	7
Poly-AC	$G_4(AC)_8G_4$	0	7
Poly-AU	$G_4(AU)_8G_4$	0	7
Poly-A	$G_4(AA)_8G_4$	0	0
Poly-AG	$G_4(AG)_8G_4$	50	0
Poly-G	$G_4(GG)_8G_4$	100	0
Arbitrary Sequences			
Label	Sequence	% GC	# YR Steps
Seq-1	G_4CCUAACAUCGAUUCGCG$_4$	50	4
Seq-2	G_4UACUGCACUAACGCGAG$_4$	50	6
Seq-3	G_4CCGGUAGCCAGGCCGUG$_4$	75	4
Seq-4	G_4AUCUUAAUGAAUCAGAG$_4$	25	3

Table 4.1: DsRNA sequences simulated in this work. The sequences are written from the 5' end to the 3' end. All nucleotides in the duplexes form canonical Watson-Crick base pairs with their complementary strand. YR-steps have been underlined in the random sequences.

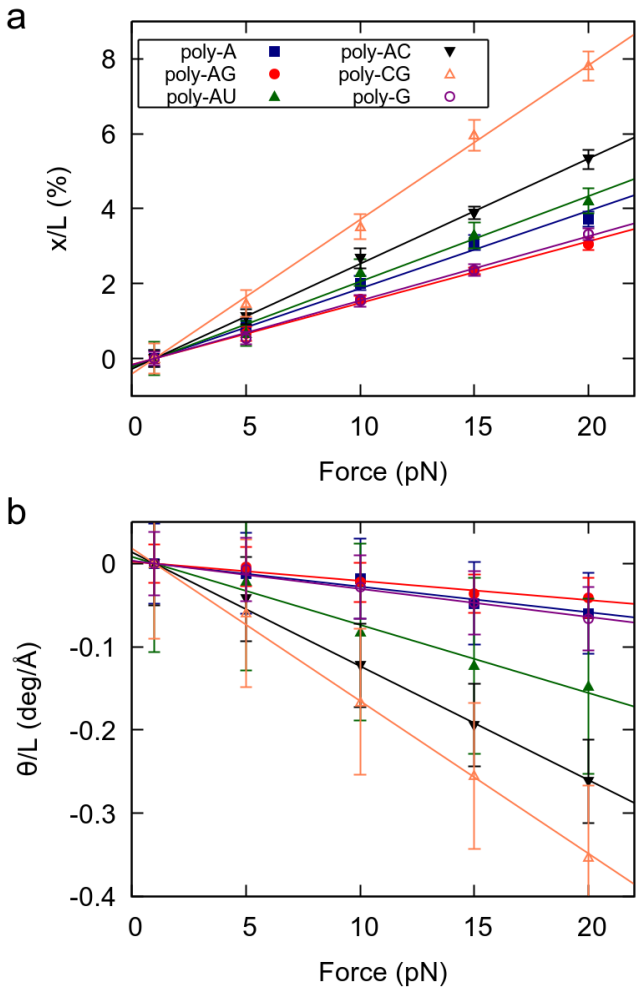

Figure 4.1: Mechanical force response of the benchmark RNA duplexes. (a) Elongation as function of the applied force. (b) Change in twist as function of the applied force. Elongation and change in twist were divided by the value of the extension at 1 pN. Color legend is the same for both panels. Data analysis was done as described in Section 1.5.1.

stiffest, all three showing an approximately similar degree of flexibility. Therefore, in terms of response to an external force, the benchmark sequences consisting of alternating purine-pyrimidine (RY) and pyrimidine-purine (YR) steps are softer than those where all steps are purine-purine (RR).

4.3 Rationalizing the flexibility of benchmark dsRNA sequences

In the previous section we unveiled a significant sequence dependence of dsRNA global mechanical properties, namely the stretching and twisting response to an external force. In the following, we aim to provide a rationalization of these findings.

4.3.1 A global approach: dsRNA crookedness fails to capture sequence effects

To gain further insight into how dsRNA sequence affects its global flexibility, we resorted to the crookedness model, developed in Section 3.3. Thus, similarly to the DNA case (Fig.3.6a), we computed the crookedness, β, and its associated flexibility, k_β, for our benchmark dsRNA molecules. This is represented in Fig. 4.2, where we included four additional dsRNA molecules of arbitrary sequence (see below and Table 4.1). Notice that, contrary to the dsDNA case, there is no one-to-one correspondence between the crookedness and dsRNA flexibility. Therefore, we will seek an explanation of dsRNA sequence-dependent flexibility by looking at the local deformability of the individual base pair steps.

4.3.2 A local approach: pyrimidine-purine steps dominate the elastic response of dsRNA

With the aim of quantifying the local flexibility of the RNA duplexes, we analyzed the force response of each base step kind. We computed the force-induced variations in helical rise and helical twist for each base pair step, which quantify the contribution of a given dinucleotide to the force response of the entire duplex. We then averaged these quantities over the steps of the same kind, *e.g.* all the CG dinucleotides, see Fig. 4.3.

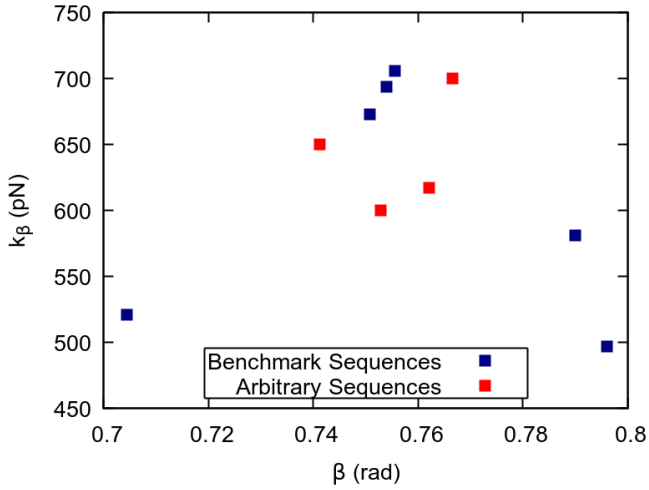

Figure 4.2: Check of the crookedness model for dsRNA. The crookedness parameter, β, and its associated flexibility, k_β were computed as in Section 3.3. Contrary to the dsDNA case (Fig. 3.6), we found no relation between β and k_β in dsRNA.

Remarkably, the helical rise and helical twist response of each step kind is linear with the force, as occurred for the whole dsRNA helices, see Fig. 4.1 and Fig. 4.3. More importantly, there is a strong variability among the base pair steps. YR steps (CG, CA and UA) are highly deformable under force, whereas RR and RY steps show a higher stiffness. Among the YR steps a high variability is found, being CG the softest followed by CA and UA in this order. On the contrary, smaller differences are observed among the rest of dinucleotides. Our results suggest that this disparate flexibility at the local level, concretely the high flexibility of the YR steps, is responsible for the differences observed in the global mechanical properties of the benchmark RNA duplexes (Fig. 4.1). Indeed, the poly-RY duplexes, for which half of the steps are of the kind YR, were systematically more flexible under an external force than their poly-R counterparts, which lack YR steps (Fig. 4.1).

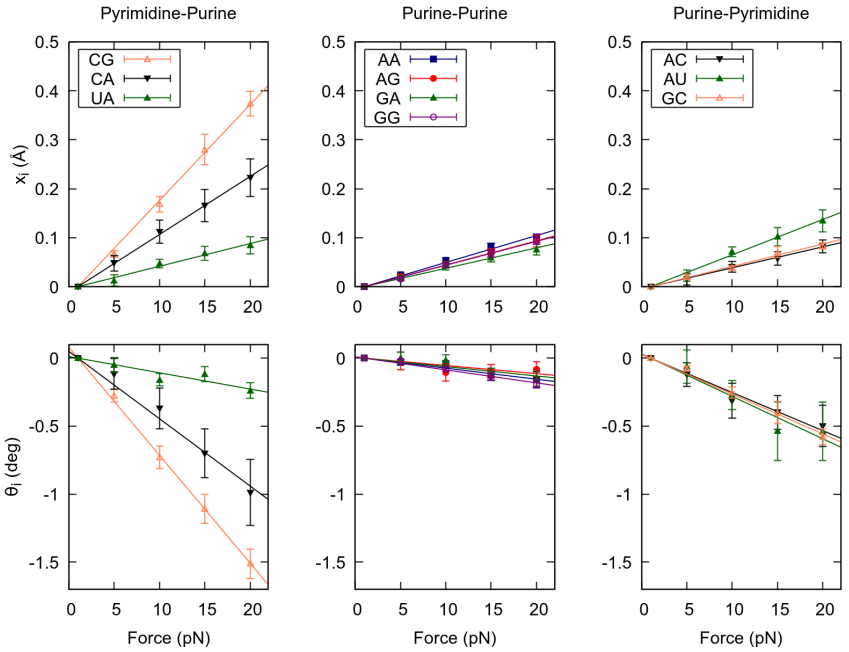

Figure 4.3: Force response of the ten kinds of base pair steps. The ten dinucleotide kinds were split into three families: pyrimidine-purine, purine-purine and purine-pyrimidine to highlight differences in flexibility between these families. The force induced change in helical rise (top panels) and helical twist (bottom panels) was computed for each dinucleotide at each external force for the benchmark sequences. These values were then averaged for all base pair steps of the same kind. To guide the eye, all data sets were fitted linear functions constrained to go through the (1,0) point. Error bars in all panels are the SEM.

4.4 The mechanical response of arbitrary dsRNA sequences

4.4.1 Pyrimidine-purine steps are highly flexible in random sequences

Our finding that pyrimidine-purine steps are highly flexible in the context of the benchmark sequences raises the question of whether this effect is also present in YR-steps of random dsRNA sequences. To address this issue, we considered a set of four randomly generated sequences with different values of fixed GC-content, see Table 4.1. We performed 1 μs-long unrestrained MD simulations of these random sequences and studied the conformational fluctuations of each base step. In parallel, we simulated the benchmark sequences in the absence of force and compared the results with the random sequences.

Consistent with their softer force response (Fig. 4.3), YR-steps showed larger helical rise and helical twist fluctuations than the rest of dinucleotides, see Fig. 4.4. These fluctuations were quantified by computing the standard deviation of these parameters for all the base pair steps and by averaging over the dinucleotides of the same kind, *i.e.* all the CG's. Importantly, we found that the flexibility trends are highly similar for both sets of sequences. This result reveals that the large flexibility of YR-steps is not exclusive to the model benchmark sequences and supports the so-called dinucleotide approximation, which assumes that the sequence in which it is embedded has a small effect on the flexibility of an individual dinucleotide. This assumption holds better for helical rise fluctuations, as quantified by the better agreement between the sets of sequences and by the shorter error bars reflecting smaller variations within a given set.

A deeper analysis of dsRNA base pair step flexibility. The covariance matrix

An extensive analysis of our dsRNA simulations in the coordinate system of base pair step parameters further supported the dinucleotide approximation, namely that dinucleotide flexibility is independent of the sequence context. The base pair step parameters is a coordinate system that describes the relative position and rotation of two base pairs in a given dinucleotide step (see Section 1.2.3, Fig. 1.3). This system comprises three translational (shift, slide and rise) and

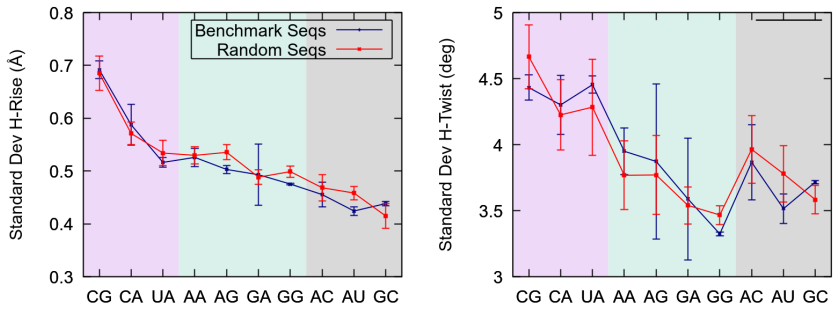

Figure 4.4: Helical rise and helical twist standard deviation of all dinucleotides. The standard deviations were computed for all the base pair steps and were then averaged for each base pair step kind. For comparison, this analysis was performed for the benchmark and random sequences separately. The shaded regions delimit the different dinucleotide families: pyrimidine-purine in pink, purine-purine in green and purine-pyrimidine in gray. A line connecting the points was included to guide the eye. Error bars in both panels are the SEM.

three rotational parameters (tilt, roll and twist). We assumed a mechanical model in which the energy required to drive a dinucleotide step away from its equilibrium conformation is harmonic, as explained in [47, 66] (see Section 1.3.2). In this case, the deformation energy can be written as

$$E = \frac{1}{2}\sum_{i=1}^{6}\sum_{j=1}^{6} k_{ij}\Delta q_i \Delta q_j, \tag{4.1}$$

where the subindices i, j refer to each of the six base pair step parameters; $\Delta q_i = (q_i - q_i^0)$ is the deviation of the i^{th} parameter from its equilibrium value; and k_{ij} are the elements of the stiffness matrix K, which quantify the energy penalty with respect to a given deformation. The elements of K can be computed by inversion of the covariance matrix $(C)_{ij} =< \Delta q_i \Delta q_j >$ as [61, 66, 163]

$$K = k_B T C^{-1} \tag{4.2}$$

where k_B is the Boltzmann constant and T the temperature of the system, which in our case is 300 K. Using Eq. 4.2, we computed the elements of the stiffness matrix for each base pair step in all the molecules and then averaged over the

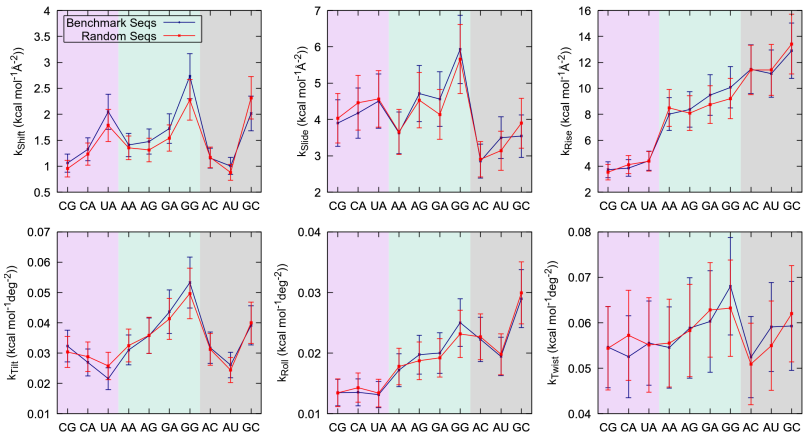

Figure 4.5: Diagonal elements of the stiffness matrix. The stiffness matrix was computed for all base pairs (using Eq. 4.2) and averaged over the steps of the same kind, *e.g.* all the CG dinucleotides. The error bars are the standard deviation of this average. The analysis was done separately for the benchmark and arbitrary sequences. The shaded regions delimit the different dinucleotide families and a connecting line was added in a similar fashion to Fig. 4.3.

same base pair step kinds, *e.g.* all the AU steps. We also computed the standard deviation of the k_{ij}'s which reflects the variation among base pair steps of the same kind located in different positions along the molecule. The diagonal matrix elements were separately analyzed for the benchmark and random sequences, see Fig. 4.5. The similar results obtained from the benchmark and random sequences support the idea presented in the main text that the dinucleotide flexibility of dsRNA is approximately independent of the global sequence context.

Finally, the covariance matrix allows obtaining the conformational volume accessible to each base pair step. This quantity provides an estimate of the general flexibility of the dinucleotide and can be computed according to [163]

$$V = det(C)^{1/2} \tag{4.3}$$

We compared the dinucleotide conformational volumes from the benchmark and random dsRNA sequences and found similar results, see Fig 4.6.

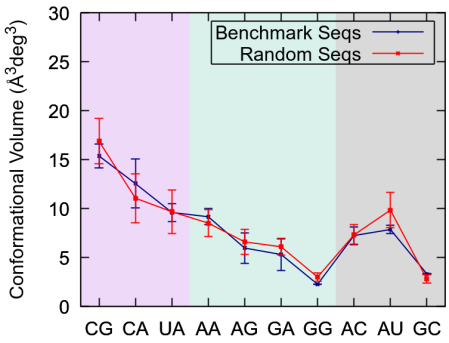

Figure 4.6: Conformational volume of the ten base pair step kinds. The conformational volumes were computed for all the base pair steps using Eq. 4.3. These values were then averaged for each base pair step kind. The error bars are the standard deviation this average. For comparison, this analysis was performed for the benchmark and random sequences separately. The shaded regions and the line connecting the points were drawn as indicated in Fig. 4.3.

4.4.2 Pyrimidine-purine steps soften the mechanical response of arbitrary dsRNA sequences

Having proved the high flexibility of YR-steps in random dsRNA sequences, we then turned our attention to the effect of this local flexibility on the global force response of the duplexes. We performed CFMD simulations of the random sequences at F =1, 5, 10, 15, 20 pN and measured the force-induced changes in extension and helical twist, as done in Fig. 4.1. Naively, one would expect that the random sequences with larger number of YR-steps will present a softer force response, as happened for the benchmark sequences. Indeed, the degree of elongation and untwisting under force reasonably correlated with the number of YR-steps of the duplexes (Fig. 4.7 and Table 4.1). Seq-2, which contains the largest number of YR steps, showed the softest response to an external force; and Seq-4, which has the fewest YR steps, was the most rigid duplex. Seq-3 and Seq-1, which have an intermediate number of YR-steps presented an intermediate force response. Based on these results we propose that the overall mechanical response of a given RNA duplex can be thoroughly tuned by

modulating the relative abundance of YR-steps in the nucleotide sequence.

4.5 Comparison of sequence effects in dsDNA and dsRNA

The comprehensive study of dsRNA flexibility presented here revealed significant differences with the sequence-dependent dsDNA mechanical properties reported in Chapter 3. Concretely, the nucleotide sequence affects the stretching response of dsDNA and dsRNA in a strikingly different manner. This can be seen in Fig. 4.8a, where we compare the values of the effective stretch modulus of our benchmark dsRNA sequences with the ones presented in Chapter 3 for their dsDNA counterparts. Notice that the poly-CG RNA duplex is exceptionally flexible, while in the DNA case experiments and simulations show that this sequence is highly stiff [56, 134, 162]; the poly-G DNA is very soft [56, 162], but one of the stiffest dsRNA sequences here studied; and the poly-A DNA is known to be extremely rigid (see Chapter 3 and 5), whereas the poly-A RNA duplex has a standard mechanical response.

This discrepancy between dsDNA and dsRNA mechanical response can be rationalized from a local approach, that is, from the dinucleotide deformability of both duplexes. We analyzed the DNA helical rise dinucleotide fluctuations of the benchmark sequences from Chapter 3 and compared the results with dsRNA (Fig. 4.8b). Although both nucleic acids share certain features, such as the exceptional flexibility of YR-steps, the patterns of dinucleotide flexibility are indeed quite distinct for dsDNA and dsRNA. For example, the AA dinucleotide is one of the stiffest dsDNA steps, but is relatively flexible in the dsRNA case; and the opposite occurs for the GG step: it is soft in the DNA duplex, but rigid in dsRNA. Importantly, other measurements of dinucleotide flexibility, namely the helical twist fluctuations and the conformational volume, also showed remarkably different sequence-dependence patterns in both nucleic acids (see Fig. 4.9). Therefore, we propose that such disparities at the local level are responsible for the differences observed in the mechanical response of the entire duplexes.

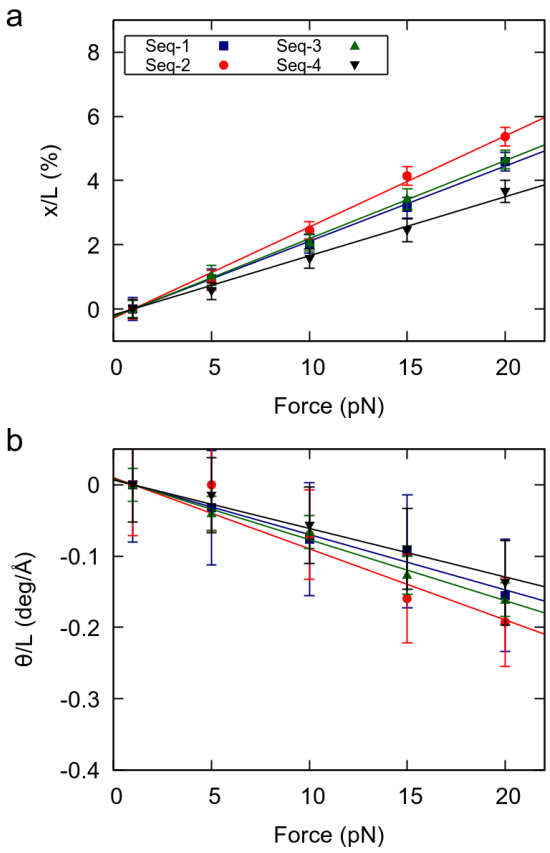

Figure 4.7: Mechanical force response of the random RNA duplexes. (a) Elongation as function of the applied force. (b) Change in twist as function of the applied force. Elongation and change in twist are divided by the value of the extension at 1 pN. Errors were obtained as described in Fig. 4.1.

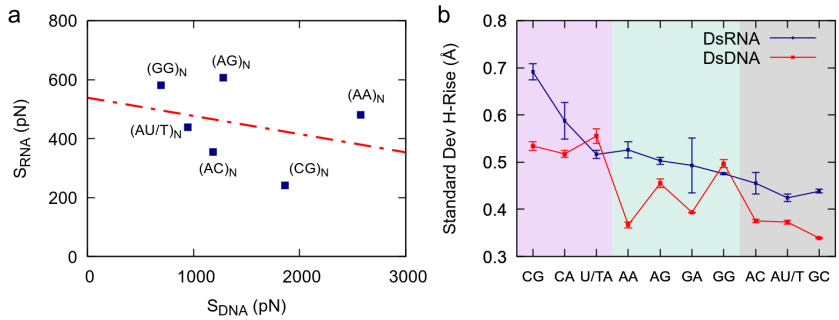

Figure 4.8: Comparison between dsDNA and dsRNA stretching flexibility. (a) The values obtained here for the stretching flexibility of the benchmark dsRNA sequences were compared with the values for the same DNA sequences (changing U by T) from Chapter 3. The effective stretch modulus of the dsRNA molecules was obtained as the inverse of the slopes of the force-extension curves (Fig. 4.3a). The number of dinucleotides comprising the sequences, N, is equal to 8 for dsRNA and 5 for dsDNA. The dsDNA and dsRNA effective stretch moduli, show a small negative Pearson correlation coefficient of r =-0.31. (b) The helical rise standard deviation of all the dsDNA and dsRNA dinucleotides was computed as described in Fig. 4.4.

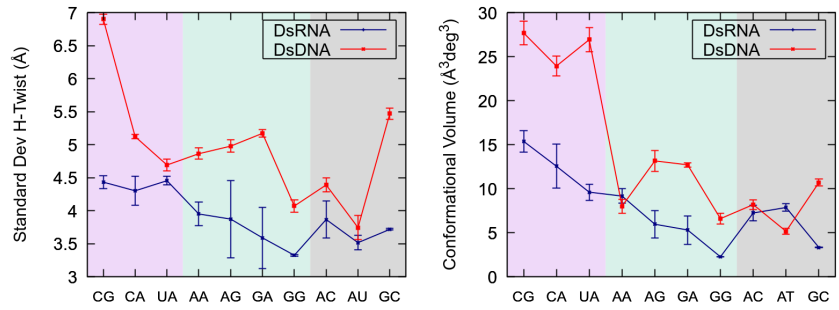

Figure 4.9: Helical twist fluctuations and conformational volumes of dsDNA and dsRNA dinucleotides. The helical twist standard deviation and the conformational volume (Eq. 4.3) was computed for all the base pair steps of the benchmark dsDNA and dsRNA sequences. We then averaged over the base pair steps of the same kind, *i.e.* all the CG steps. Error bars are the standard error of this average. The shaded regions indicate the different dinucleotide families: pyrimidine-purine (pink), purine-purine (green) and purine-pyrimidine (gray). A line connecting the points was included to guide the eye.

4.6 Discussion

4.6.1 A novel difference in the mechanics of dsDNA and dsRNA

Despite their similar chemical composition, dsDNA and dsRNA have been recently shown to exhibit two remarkable differences in their elastic response [32, 37]. Firstly, dsDNA is around three times stiffer than dsRNA with respect to stretching deformations [32, 34, 36]. Secondly both duplexes possess an opposite twist-stretch coupling [37, 42, 43]. Here we propose a third fundamental difference between dsDNA and dsRNA: the role of the nucleotide sequence on the overall flexibility of the duplex.

At first sight, this result might seem controversial: one may argue that because the nucleotide bases are the same in both nucleic acids - excepting uracils/thymine -, their effect on the flexibility of the double-helix should be similar. However, this argument can be rejected in simple terms by resorting to the geometry of the DNA and RNA double-helices, namely the A- and B-forms. In the A-form, consecutive base pairs are highly inclined with respect to each other, whereas in the B-form they are approximately parallel. As a result of these disparate geometries, the chemical interactions between the stacked base pairs are strongly affected, altering their flexibility. For example, the conformations of a given base pair in a highly inclined A-form configuration might be constrained due to steric clashes with its neighbor. However these steric clashes might not be present in B-form planar stacking, allowing the dinucleotide to explore a broader conformational space. In this hypothetical case, the dinucleotide would be rigid when found in dsRNA but flexible in dsDNA. Furthermore, even if the local dinucleotide deformability were similar for both molecules, the projection of these deformations on the helical axis would likely differ in the two helical geometries. This idea is introduced in [65]. In the case of an RNA A-form helix with inclined base pairs, sliding deformations parallel to the base pairing can substantially elongate the duplex. On the contrary, these very same base pair step deformations can barely stretch the B-form helix, where these sliding movements are practically orthogonal to the helical axis. Consequently, a dinucleotide with high slide flexibility is likely to play a more important role in the

force response of dsRNA when compared to dsDNA.

4.7 Conclusions

We performed CFMD simulations on dsRNA duplexes and found that their global mechanical properties are strongly sequence-dependent. This finding was rationalized from a local perspective: the enhanced flexibility of pyrimidine-purine dinucleotides plays a critical role in the global deformability of the duplex. Consistently, when inserted in dsRNA molecules of random sequences, the number of YR-steps correlated with the overall flexibility of the helix. Our results pave the way towards the design of dsRNA sequences with predefined mechanical properties for biophysical and nanotechnology applications.

5 Sequence induced bending in dsDNA: A-tracts

Adapted from: A. Marin-Gonzalez, C. L. Pastrana, R. Bocanegra, A. Martin-Gonzalez, J. G. Vilhena, R. Perez, B. Ibarra, C. Aicart-Ramos, F. Moreno-Herrero, *Understanding the paradoxical mechanical response of in-phase A-tracts at different force regimes*, bioRxiv, doi:10.1101/854968

5.1 Introduction

A-tracts are DNA sequences consisting of four or more consecutive A:T base pairs without the TA step. They are widespread across the genomes of both prokaryotic and eukaryotic organisms, including humans [164–167]. Notably, the distribution of A-tracts along their genomes has proven to be essential for the proper organization of their genetic material with implications in transcription regulation [121, 122, 144, 168, 169]. In addition, A-tracts have been shown to play an important role in recombination [119, 170], replication [171], antiviral response [13, 172] and stochastic gene silencing [173].

Remarkably, many of the functions of A-tracts have been linked to their particular structure and mechanical properties [115, 119, 122, 170]. Regarding the structure, A-tracts are known to introduce a directional bend in the DNA helical axis [119]. When two or more A-tracts are located in phase with the helical pitch, their bending adds constructively leading to a significant global curvature of the molecule [26]. This curvature contrasts with the anomalously straight conformation reported for a single A-tract [25, 174]. Possibly, the most widely accepted solution to this conflict is the junction model, in which the bending is primarily localized at the edges of the A-tracts [26]. Nevertheless,

the precise bending mechanism in A-tracts remains a matter of debate [119, 175, 176].

In parallel, the reported mechanical properties of A-tracts are to some extent paradoxical. Early crystallographic studies suggested that A-tracts are stiff [25, 177, 178]. This rigidity is in line with the high stretch modulus reported by recent molecular dynamics simulations of short (∼15 bp) duplexes containing an A-tract [62, 162], and can be attributed to the distinct structure of these sequences, namely their low crookedness (see Chapter 3). On the other hand, cyclization studies performed on duplexes longer than ∼100 bp revealed that A-tracts are not particularly rigid [179, 180]. Other studies revealed that they are even more flexible than standard DNA sequences [181]. This is supported by single-molecule experiments showing an enhanced curvature and looping probability in A-tracts [89, 182, 183]. However, at the intermediate scale of ∼70 bp, insertion of a poly-A inside a random sequence greatly reduces its looping rate [147], a sign of high rigidity.

Taken together, these results reveal a remarkable length-dependence of the mechanical properties of A-tracts and call for a unified comprehensive study. Such description should rationalize why these molecules appear rigid at scales of one helical turn, but flexible at scales longer than ∼100 bp. Moreover, a full characterization of the mechanical properties of A-tracts should quantitatively distinguish their entropic bendability from their intrinsic static bending. This task is highly non-trivial in either structural or cyclization studies and relies on precise knowledge of the trajectories of the molecules over distances of hundreds of base pairs.

In this Chapter, we study the mechanical properties of phased A-tracts at multiple forces and length scales, and at the single-molecule level. Atomic force microscopy (AFM) imaging shows that phased A-tracts induce long-range bending on DNA molecules. The bending is explained due to the presence of an intrinsically bent structure and is rationalized using a newly developed theoretical model[1]. Interestingly, phased A-tracts significantly soften the mechanical

[1]When we finished building our model (September, 2018), an article was published reporting a similar theoretical approach to describe the mechanical properties of collagen fibers

response of DNA molecules under low stretching forces. Nevertheless, at high forces the situation is reversed and the A-tracts confer DNA an unprecedented rigidity. Altogether, our work unveils the complex interplay between structural and mechanical properties of A-tracts across multiple force scales.

5.2 Mechanical properties of A-tracts in the absence of external force

5.2.1 In-phase A-tracts induce intrinsic bending over long distances

We fabricated long DNA molecules suitable for Atomic Force Microscopy (AFM) imaging containing phased A-tracts from an intronic fragment of the *C. elegans* genome, hereafter the intron (see Fig. 5.1a). This construct contained three copies of the intron, resulting in a 2636 bp DNA with phased A-tracts. As control, we considered a DNA molecule of approximately the same length and with a GC content centered around $\sim$50% (see Fig. 5.1b). As shown in Fig. 5.2, the A-tracts molecule presented a slower gel migration than the control, a classical signature of intrinsic bending in DNA [26]. This was also observed for the DNA constructs fabricated for Magnetic and Optical Tweezers experiments (see Sections 5.3, 5.4). Synthesis of DNA molecules and gel electrophoresis experiments were performed by Clara Aicart-Ramos from Fernando Moreno-Herrero lab.

AFM has proven to be an invaluable tool for characterizing DNA mechanical properties, providing a rigorous experimental test of DNA polymer models [30–32, 182]. Following the conditions described in Section 1.5.2, we used an AFM to capture two-dimensional equilibrium conformations of random (control, Fig. 5.3a) and phased A-tracts DNA molecules (Fig. 5.3b). As a first quantitative characterization, we computed the mean contour length of the control and the A-tracts molecules, obtaining values of 897 $\pm$ 17 nm (n=25, error is std) and 892 $\pm$ 14 nm (n=23), respectively. These values are very similar to each other and both yield a helical rise per base pair of 0.34 nm bp^{-1}, in agreement with the standard value of B-DNA [17].

By visual inspection one can already notice that the A-tracts molecules ap-

[184].

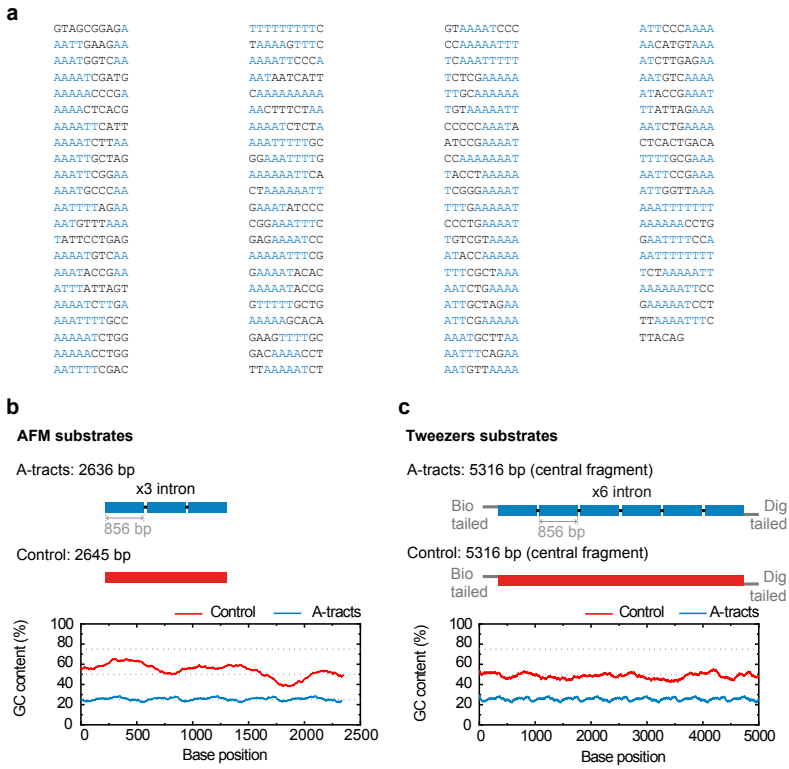

Figure 5.1: Molecules used in this study. (a) Intron sequence studied in this work as reported in [89]. A-tracts (regions of four or more consecutive A's and T's without the TA step) have been marked in blue. The sequence was written in columns of 10 letters to highlight the $\sim$ 10 bp periodicity. (b) Top, the A-tract AFM substrate consisted on 3 repetitions of the intron, depicted by small blue blocks. Shown in red, the control molecule of approximately the same length. Bottom, the GC-content of the A-tract and the control construct. (c) Top, the MT and OT molecules contained six repetitions of the intron and were flanked by two oligonucleotides labeled with biotin and digoxigenin. The control construct has the same length as the six repetitions of the intron. Bottom, GC-content of A-tract and control tweezers constructs.

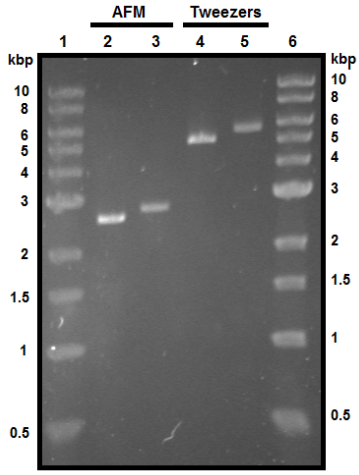

Figure 5.2: Migration of control and A-tracts constructs in agarose gels. 0.8 % agarose gel electrophoresis of AFM substrates (control DNA (2645 bp) lane 2 and A-tracts (2636 bp) lane 3), and central insert of the magnetic and optical tweezers substrates (control, lane 4 (5272 bp) and A-tracts, lane 5) were run 30 ng/lane). Lanes 1 and 6 correspond to 1 kbp DNA ladder (NEB). The electrophoresis was run at room temperature for 50 min at 90 V and the gel was later stained with SYBR®Safe (Invitrogen). The A-tracts constructs migrated slower than the controls, despite having the same length.

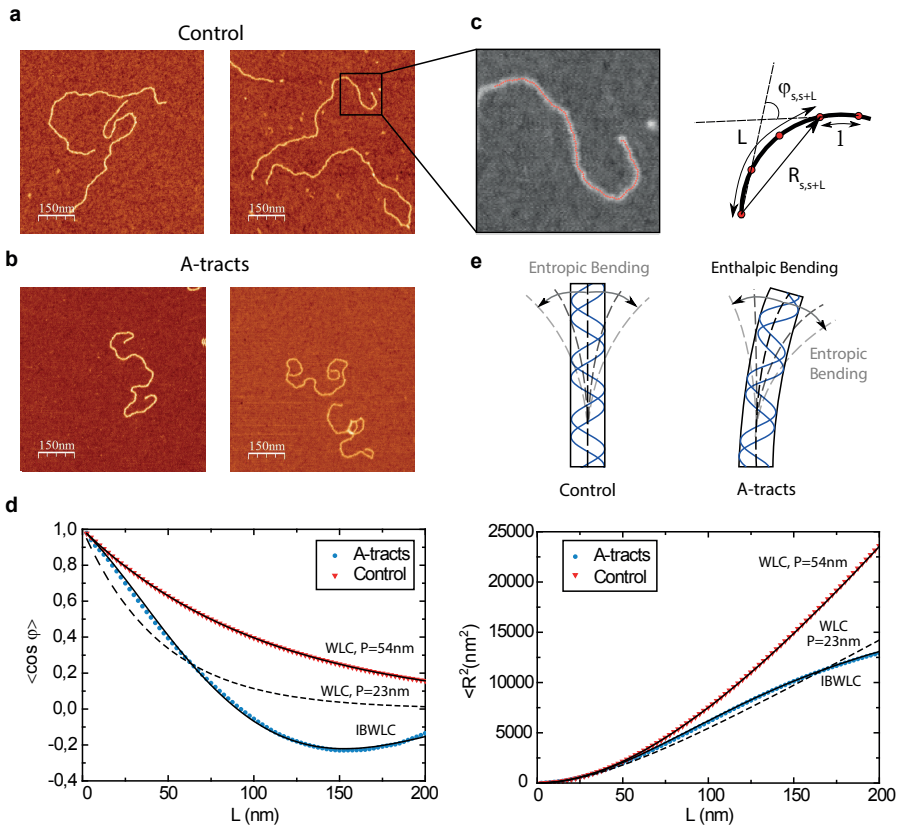

Figure 5.3: Mechanical properties of A-tracts at zero force. Representative AFM images of (a) the Control and (b) the A-tract DNA molecules. (c) Illustration of a typical trace. Left, zoom of the marked region with a trace of points separated by 2.5 nm that follow the contour of the DNA molecule. Right, schematic representation of the tangents, $\cos\varphi_{s,s+L}$ and the distance, $R_{s,s+L}$ for two points separated by a contour distance of $L = 7.5$ nm. (d) $\cos\varphi$ and R^2 were averaged over all the traces and plotted as a function of the contour distance separation, L. Left, the cosine data of the control and A-tract molecules were fitted to the WLC and IBWLC using Eq. 5.2 and Eq. 5.5, respectively. This yielded a value of $P = 54 \pm 1$ nm for the control; and values of $P = 55 \pm 1$ nm, $a = 17.4 \pm 0.1$ μm^{-1} for the A-tract molecule. Right, these values of the parameters were used to plot the theoretical expression of $< R^2 >$ in the WLC and IBWLC (Eq. 5.3 and Eq. 5.6, respectively). In addition, the A-tracts data was fitted to the WLC model (dashed line) yielding $P = 23 \pm 1$ nm. (e) Schematic representation of the WLC and IBWLC models.

pear more bent than the control ones, in line with the findings reported in [89]. In order to quantify the flexibility of these molecules we firstly computed their traces following a previous protocol [32, 89]. We calculated pair of coordinates separated by a distance of $l = 2.5$ nm, that follow the contour described by the adsorbed DNA molecules (Fig. 5.3c). From these traces, we obtained the angle, $\varphi_{s,s+L}$, defined by the tangents at two points separated by a contour distance L; and the distance, $R^2_{s,s+L}$, between these points (Fig. 5.3c). We then averaged $R^2_{s,s+L}$ and $\cos\varphi s, s+L$ over all the points of the trace and over hundreds of traces. Representing these two quantities, $< R^2_{s,s+L} >$ and , $< \cos\varphi_{s,s+L} >$ as a function of L allows direct comparison of AFM data with polymer models (Fig. 5.3d).

The mechanical properties of DNA molecules are usually analyzed in the context of the worm-like-chain (WLC) model, which describes polymers with harmonic bending energy

$$E = \frac{Pk_BT\varphi^2}{2L}, \tag{5.1}$$

where φ is the bending angle; L the contour separation as defined for the AFM traces; and P is the persistence length, which is proportional to the bending stiffness: $P = B/k_BT$ [31]. Notice that the minimum energy configuration corresponds to $\varphi = 0$, that is, to straight molecules. Therefore, according to the WLC model, DNA molecules are bent solely by thermal fluctuations, *i.e.*, they are entropically bent. The AFM data can then be fitted to the equations of the WLC in two dimensions:

$$< \cos\varphi >= e^{-L/2P} \tag{5.2}$$

$$< R^2_{s,s+L} >= 4P\left(L + 2P\left(e^{-L/2P} - 1\right)\right) \tag{5.3}$$

In agreement with previous works, our control data nicely fit to Eq. 5.2 and Eq. 5.3 with $P = 54 \pm 1$ nm [31, 32] (Fig. 5.3d). However, A-tracts data deviated from the WLC behavior, and the best fit to Eq. 5.3 provided an extremely low persistence length of $P = 23 \pm 1$ nm (Fig. 5.3d, right panel). This deviation is even more remarkable for the cosine's correlation, for which the WLC predicts an exponential decay (Eq. 5.2). In contrast, the cosine's correlation of the A-tracts molecules reached negative values, with a minimum of $< \cos\varphi >\sim$-0.2

at around $L = 150$ nm (Fig. 5.3d, left panel). This demonstrated that, at zero force, the simple picture of the WLC model that assumes purely entropically bent polymers is not sufficient to describe the flexibility of A-tracts DNA molecules.

5.2.2 An intrinsically-bent worm-like chain model captures the long-range bending induced by in-phase A-tracts

We solved this discrepancy by building a theoretical model that describes polymers that are not only entropically, but also enthalpically (or intrinsically) bent. This model was accordingly called intrinsically-bent WLC (IBWLC) and was inspired by the works of Rivetti et al [31, 182]. Remarkably, our IBWLC model has a simple, analytical solution that can be easily fitted to the AFM data, allowing to quantitatively decouple the bending contribution arising from thermal fluctuations (entropic contribution) from that coming from the intrinsic structure of the A-tracts (enthalpic contribution). This model is built from basic physical assumptions and thus, is not only applicable to DNA molecules, but to any other polymer showing such interplay between entropic and enthalpic bending deformations. This model is based on the assumption that the minimum of bending energy corresponds to a bent trajectory described by the arc of a circle of radius R_0. This assumption has been widely used in the literature to explain that A-tracts induce a bend in the structure of the DNA [119], as well as cyclization [176, 185] and AFM data [89]. This assumption is sketched in Fig. 5.3e and can be expressed mathematically as

$$E = \frac{P_0 k_B T(\varphi - aL)^2}{2L} \tag{5.4}$$

where $a = 1/R_0$ is the intrinsic curvature, and P_0 will be termed as natural persistence length of the molecule, which quantifies its resistance to bending similar to the WLC case. Notice that the WLC expression (Eq. 5.1) is recovered in a straightforward manner by making $a = 0$, that is, by deleting the intrinsic curvature; and by substituting P_0 by P. From Eq. 5.4 we obtained the relevant $< \cos\varphi_{s,s+L} >$ and $< R^2_{s,s+L} >$ expressions for fitting the experimental data to the IBWLC, in a similar fashion to the WLC case [31]:

$$< \cos\varphi_{s,s+L} >= e^{-L/2P_0}\cos(aL) \tag{5.5}$$

$$< R^2 > = \frac{2}{a^2+b^2} \times \\ \times \left\{ bL + \frac{1}{a^2+b^2} \left[(a^2-b^2)(1-\cos(aL)e^{-bL}) - 2ab\sin(aL)e^{-bL} \right] \right\} \tag{5.6}$$

where we have defined $b \equiv 1/(2P_0)$ for convenience. A full derivation of the model can be found in Appendix D.

A fit of the IBWLC model (Eq. 5.5) to the A-tracts $< \cos\varphi >$ data provided the fitting parameters $a = 17.4 \pm 0.1$ μm^{-1} and $P = 55 \pm 1$ nm (Fig. 5.3d, left panel). Using these values, we plotted the theoretical expression of $< R^2_{s,s+L} >$ (Eq. 5.6) and found an excellent agreement with the experimental data (Fig. 5.3d, right panel). Remarkably, this agreement held for contour distances that extended up to at least 200 nm.

The IBWLC model enabled us to quantitatively dissociate the entropic and enthalpic contributions to the A-tracts flexibility. The value that we found for the intrinsic curvature was close to a previous estimation of ~ 10 μm^{-1} obtained from AFM images [89]; and the persistence length essentially coincided with the control and with indirect measurements from cyclization studies on A-tracts [179, 180]. In addition, our results were in good agreement with previous AFM experiments on DNA molecules containing few (less than eight) A-tracts [182]. Altogether, our results indicate that the A-tracts molecule has no enhanced bending flexibility. In other words, deviations from the minimum energy are energetically as costly as for the control molecule, as quantified by a similar persistence length for both molecules. On the contrary, the apparent larger bendability of this molecule stems from a purely structural or static deformation, its intrinsic bending.

5.3 The low-force response of A-tracts deviates from entropic models

AFM experiments show that the intrinsic bending of A-tracts is responsible for the anomalous mechanical properties of these molecules at zero force. However, inside the cell the DNA is often subjected to mechanical stress, and to

which extent the intrinsic bending of DNA affects its extension under mechanical force is unclear. In order to explore the mechanical response of phased A-tracts sequences MT were employed to obtain force-extension curves at standard 100 mM NaCl in a regime of forces <6 pN. MT experiments were performed and analyzed by Cesar L. Pastrana, from Fernando Moreno-Herrero lab. Averaged force-extension curves of A-tracts molecules reported lower extensions compared to the control data at all forces, such that the curve appears shifted to the left (Fig. 5.4a). Importantly, this decrease in the end-to-end distance of the molecules was also found at zero force in the AFM experiments and was rationalized to be an effect of the bends (Eq. 5.6 and Fig. 5.3d, right panel). Notably, this effect virtually disappears at a force of $\sim$ 5 pN, where both molecules show a similar extension and practically reach their full contour length ($\sim$ 1.8 μm).

In order to quantify the mechanical properties of A-tracts we resorted to the WLC interpolation formula [186]:

$$F(z) = \frac{k_B T}{P}\left[\frac{1}{4\left(1-\frac{z}{L}\right)^2} - \frac{1}{4} + \frac{z}{L} + \sum_{i=2}^{i\leq 7} a_i \left(\frac{z}{L}\right)^i\right] \tag{5.7}$$

where z is the extension of the molecule, F the applied force, P the persistence length and the phenomenological constants a_i are a_2 = -0.5164228, a_3 = -2.737418, a_4 = 16.07497, a_5 = -38.87607, a_6 = 39.49944 and a_7 = -14.17718.

As expected, force extension curves of control DNA were well described by the WLC model (Fig. 5.4a, red symbols) and yielded a persistence length of P = 44 $\pm$ 4 nm, in agreement with previous measurements [32, 37]. However, the WLC model could not properly fit the A-tracts data over the entire force range. At forces below 1 pN the data were best described by a P = 15 $\pm$ 1 nm, whereas in the 1-4 pN force range data were better fitted to a higher persistence length of P = 21 $\pm$ 8 nm. The deviation of the A-tracts data from the WLC model likely reflects the presence of intrinsic bends in the DNA, given that the model describes polymers that are purely entropically bent. Consequently, we propose that, similar to the zero force case (Fig. 5.3d), the deviation from the WLC together with the anomalously small persistence length constitute a signature of the presence of intrinsic bends. Therefore, in absence of a better model, we

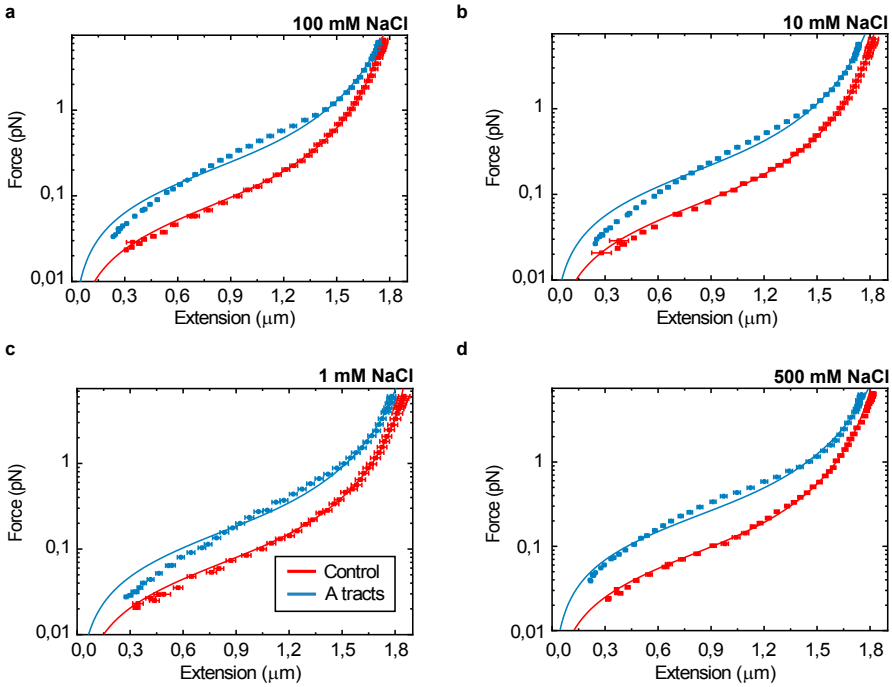

Figure 5.4: Average MT force-extension curves of A-tracts (blue) and control (red) molecules at different NaCl concentrations. Error bars are sem. The data were fitted to Eq. 5.7. The values of the fitting parameters are represented in Fig. 5.5.

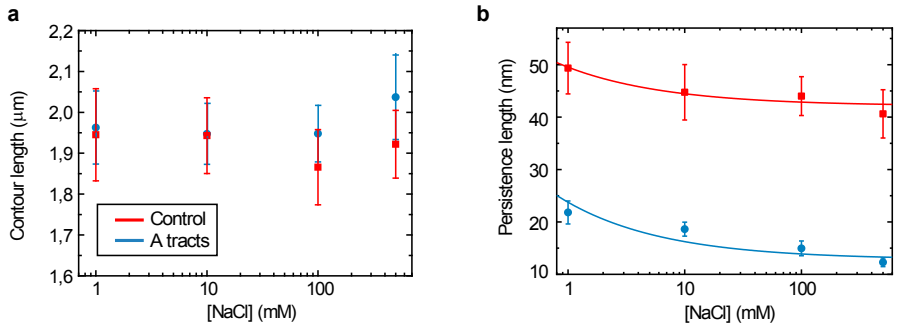

Figure 5.5: A-tracts mechanical parameters obtained from MT experiments. Contour (a) and persistence length (b) of the control (red) and A-tracts (blue) molecules at different NaCl concentrations were obtained by fitting the force-extension curves from Fig. 5.4 to Eq. 5.7. The data from (b) was fitted to $P = P_\infty + mc^{-1/2}$, where c is the salt concentration [187]. We found $P_\infty = 42 \pm 1$ nm, $m = 7 \pm 2$ nm·mM$^{1/2}$ for control molecules and $P_\infty = 14 \pm 1$ nm, $m = 11 \pm 4$ nm·mM$^{1/2}$ for A-tracts, indicating that A-tract intrinsic curvature is not strongly dependent on NaCl concentration.

will use here the WLC and quantify the bent character of the A-tracts molecules from the low value of their persistence length.

It has been argued that A-tracts bending may be the consequence of the interaction of monovalent ions with the minor groove [188, 189]. Conversely, other works favor an intrinsic ion-independent bending model as a result of the distinct base-pairing and stacking of AT-rich domains [190, 191] and even suggest a loss of curvature with the increase of monovalent salt concentration [192]. In order to elucidate the role of monovalent ions on A-tracts bending, we obtained additional force-extension curves at NaCl concentrations of 1, 10 and 500 mM (Fig. 5.4b-d) and compared them with control sequences. Fits from the WLC model resulted in similar values of contour lengths of both molecules and for all salt concentrations (Fig. 5.5a). In addition, we found that the persistence length of the A-tracts and the control molecules followed a similar decaying trend with increasing NaCl concentrations, as it has been reported for DNA molecules of arbitrary sequence [32, 36] (Fig. 5.5b). However, in all cases, the WLC could

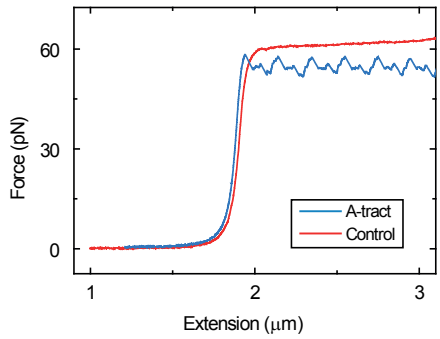

Figure 5.6: Representative OT force-extension curves showing the overstretching transitions of control (red) and A-tracts (blue) at forces ∼60 pN.

not fit the force-extension data of A-tracts sequences (Fig. 5.4). Therefore, we conclude that A-tracts bending is intrinsic to the structure of the molecule and does not rely on monovalent ions.

5.4 A-tracts present a high stretching rigidity

Further to our analysis in the entropic low force regime, we then investigated the mechanical response of A-tracts in the enthalpic regime by determining their stretching rigidity. We used optical tweezers (OT) to obtain force-extension curves of the control and the A-tract molecules in a buffer containing 100 mM NaCl. OT measurements were done by Dr. Rebeca Bocanegra from Dr. Borja Ibarra's lab. As expected for random DNA sequences, we observed a smooth overstretching transition at $\sim$ 60 pN for the control molecule [34, 90, 193] (Fig. 5.6). However, the A-tract molecule overstretching started at a lower force, showing characteristic periodic saw-tooth pattern with six repetitions. These repetitions likely correspond to the six copies of the intron (Fig. 5.1c), in line with previous OT experiments reporting reproducible sequence-dependent unpeeling of overstretched DNA [45].

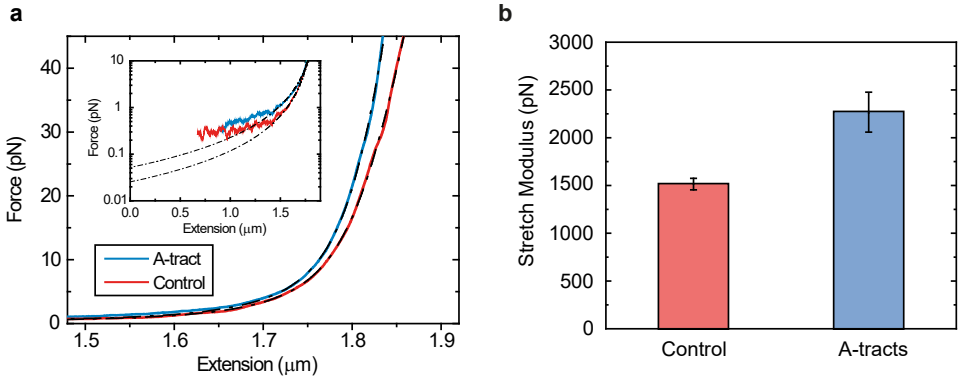

Figure 5.7: Mechanical response of A-tracts to forces above 10 pN. (a) Representative force-extension curves of control and A-tract molecules measured with OT. Dashed lines represent the fits to the eWLC (Eq. 5.9) of the control and A-tract molecules in the 10-45 pN range. The values of these particular fits were L = 1852 nm, P = 44 nm, S = 1704 and L= 1851 nm, P = 39 nm, S = 2802 nm for the control and A-tract molecules respectively. Note that the eWLC does not fit well the A-tracts data in the 1-10 pN interval. Inset, the same experimental curves are represented using a logarithmic force scale and showing the fits to the WLC in the 1-10 pN range. The values of these fits were L=1863 nm, P =41 nm and L=1916 nm, P =20 nm for the control and A-tracts data respectively. (b) Average stretch modulus (S) of control and A-tracts molecules (see Table 5.1). Error bars represent the standard error of the mean of the fits.

To quantify the stretching response of A-tracts molecules, we first considered the intermediate force regime (F<10pN) and fit force-extension data to the inextensible WLC model (Fig. 5.7b) [32]:

$$z = L\left(1 - \frac{1}{2}\sqrt{\frac{k_B T}{FP}}\right) \tag{5.8}$$

The values obtained for the control and the A-tracts molecule were $L = 1844 \pm 3$ nm, $P = 40 \pm 2$ nm and $L = 1856 \pm 9$ nm, $P = 19 \pm 1$ nm, respectively, in agreement with the AFM and MT results (Fig. 5.3d and Table 5.1). In addition, similar to the MT experiments, the A-tracts data showed deviations from the WLC fit (Fig. 5.4).

With the aim of characterizing the A-tracts mechanical properties at high forces we then resorted to the extensible WLC (eWLC) model, which considers that the molecule can be elongated beyond its contour length [36, 194]. According to the eWLC, the extension of a DNA molecule depends on the applied force as:

$$z = L\left(1 - \frac{1}{2}\sqrt{\frac{k_B T}{FP}} + \frac{F}{S}\right) \tag{5.9}$$

where S is the stretch modulus and quantifies the enthalpic elongation of the molecule. We used this expression to fit the force-extension curves of the control molecule in the 10-45 pN range. Remarkably, in this range of forces, Eq. 5.9 fits well the experimental data (Fig. 5.7c). We obtained $L = 1825 \pm 3$ nm, $P = 46 \pm 2$ nm and $S = 1550 \pm 70$ pN for the control molecule, in agreement with the exhaustive literature on DNA flexibility [32, 45, 46] (Table 5.1). The values of the fitting parameters obtained for the A-tracts molecule were $L = 1794 \pm 8$ nm, $P = 43 \pm 3$ nm, $S = 2270 \pm 170$ pN and revealed two notable features. Firstly, at high forces the entropic flexibility of the A-tracts substrate was described by a persistence length of 43 nm, similar to that of the control molecule. Secondly, the A-tracts enthalpic rigidity was extremely high, with a value of the stretch modulus as large as 2270 pN (Fig. 5.7d). Consistently, similar results were obtained at NaCl concentrations of 50 mM and 500 mM (Table 5.1).

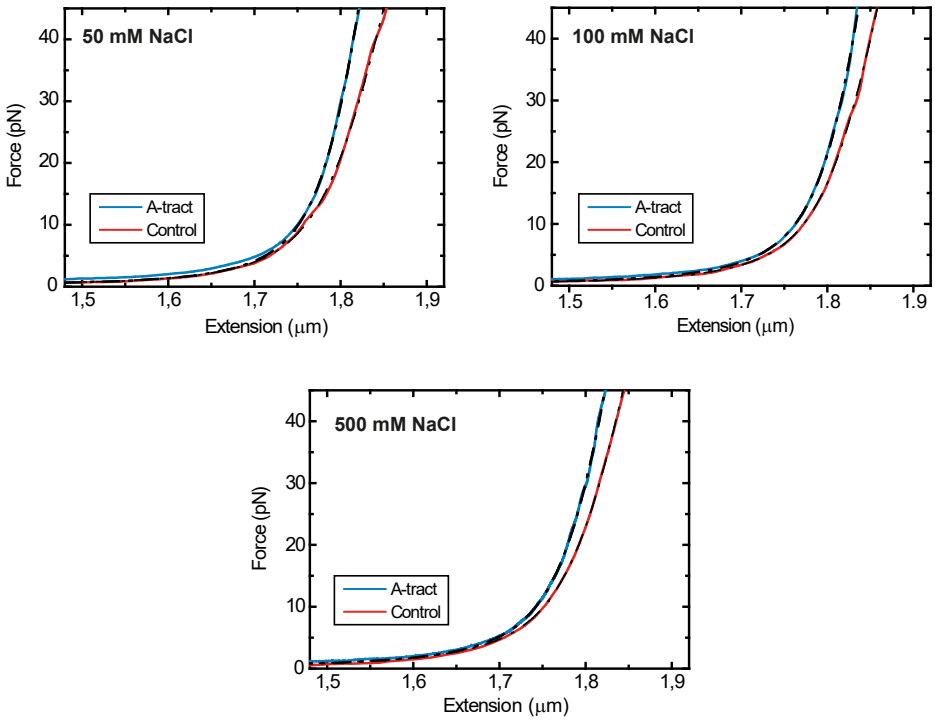

Figure 5.8: Representative optical tweezers force-extension curves of A-tracts (blue) and control (red) molecules at different NaCl concentrations. All data sets were fitted to the eWLC (Eq. 5.9) in the 10-45 pN force range. The data and the fit at 100 mM NaCl is the same as in Fig. 5.7a. The values of the fits shown at 50 mM NaCl were L= 1834 nm, P = 45 nm, S = 1440 pN and L= 1828 nm, P = 48 nm, S = 2545 pN for the control and A-tracts, respectively; and the ones at 500 mM NaCl were L= 1841 nm, P = 34 nm, S = 1675 pN and L = 1833 nm, P = 35 nm, S = 2286 pN for the control and for the A-tracts.

Magnetic Tweezers					
[NaCl] (mM)	Molecule	L (nm)	P_{WLC} (nm)	P_{eWLC}	S (pN)
1	A-tract	1936 ± 30	22 ± 1		
	Control	2013 ± 35	49 ± 1		
10	A-tract	1928 ± 20	19 ± 1		
	Control	1964 ± 40	45 ± 1		
100	A-tract	1913 ± 10	15 ± 1		
	Control	1846 ± 25	44 ± 1		
500	A-tract	1930 ± 20	12 ± 1		
	Control	1970 ± 20	41 ± 1		
Optical Tweezers					
[NaCl] (mM)	Molecule	L (nm)	P_{WLC} (nm)	P_{eWLC}	S (pN)
50	A-tract	1891 ± 5	19 ± 1	40 ± 3	2320 ± 170
	Control	1829 ± 4	44 ± 2	55 ± 3	1640 ± 110
100	A-tract	1856 ± 9	19 ± 1	43 ± 3	2270 ± 170
	Control	1844 ± 3	40 ± 2	46 ± 2	1550 ± 70
500	A-tract	1843 ± 3	22 ± 1	38 ± 3	2540 ± 180
	Control	1832 ± 3	39 ± 2	49 ± 3	1560 ± 70

Table 5.1: Mechanical properties of A-tracts and Control molecules obtained from MT and OT experiments. The parameters shown are the average of the fits (N>10 molecules) and the errors are the sem. L and P_{WLC} were obtained by fitting the MT (OT) data to Eq. 5.7 (Eq. 5.8) in the $F < 1$ pN ($F < 10$ pN) range. P_{eWLC} and S were calculated by fitting the OT data to Eq. 5.9 in the 10-45 pN force range.

5.5 Discussion

5.5.1 A comprehensive picture of A-tracts mechanical properties

Our AFM, MT and OT results can be brought together to build a comprehensive picture of the multiscale flexibility of A-tracts (Fig. 5.9). At zero force, the in-phase bends of the A-tracts result in a highly curved molecule, as observed in the AFM images (Fig. 5.9a). These bends are responsible for the measurement of a distinctively small persistence length in the A-tracts molecules, as found in MT and OT experiments at low forces. We propose that under an external force, these bends act as hinges and are gradually opened, allowing the molecule to eventually extend to its full contour length (Fig. 5.9b). Remarkably, such process of straightening the bends would represent an additional source of flexibility that is not captured by the WLC, which only considers entropic elongation. This would explain the deviations from the WLC observed in force-extension curves in the middle-force regime (F <10 pN) (Fig. 5.4). At high forces (F >10 pN) A-tracts bends would be straightened (Fig. 5.9c) and both the control and the A-tracts molecule show similar values of extension and persistence length (Fig. 5.4, 5.7 and Table 5.1). Interestingly, a similar idea has been proposed in [195] in the context of the so-called kinkable WLC model. In that work, it is proposed theoretically that a kinked elastic polymer under force would present a small effective persistence length at low forces and would recover its natural persistence length in the high force limit.

According to our model, at high forces (F>10 pN), the A-tracts molecules would be practically straightened and the local stretching rigidity of the A-tracts evaluated (Fig. 5.9c). We found an unusually high value for the A-tracts stretch modulus (Fig. 5.7d), in agreement with molecular dynamics simulations of $\sim$ 15 bp-long poly-A molecules (see [56, 62, 162] and Fig. 3.6a). Moreover, this finding confirms the high rigidity usually attributed to these sequences in structural studies [25, 177, 178]. A possible explanation for the high A-tract stiffness was proposed in Chapter 3 in the context of the crookedness model. Inside the A-tract the base pair centers are almost perfectly aligned with the helical axis and thus, the molecule can only elongate by unstacking its base

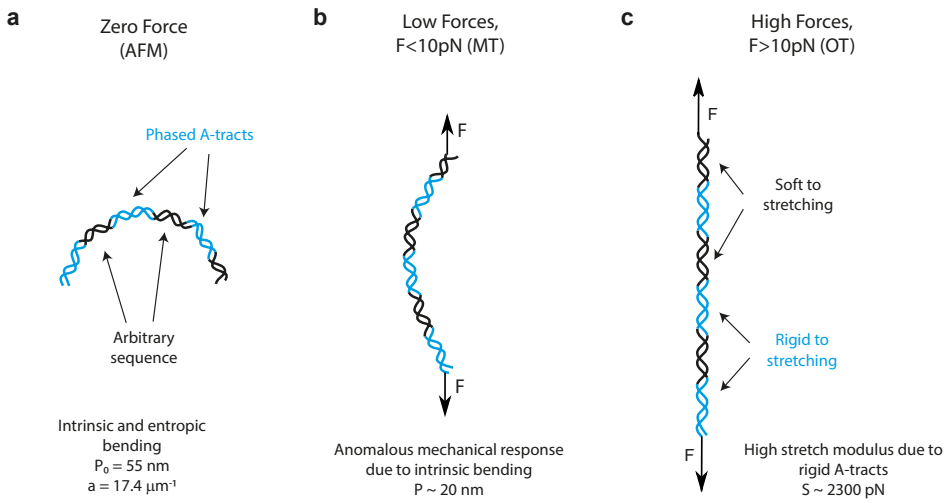

Figure 5.9: Schematic representation of the proposed model of the mechanical properties of A-tracts. (a) In the absence of stretching forces, molecules containing phased A-tracts (blue regions) appear largely bent as a consequence of the local structural bends. (b) Forces <10 pN align and straighten the bends along the pulling coordinate, which results in a significant deviation of the force-extension curves from the WLC model. (c) At forces >10 pN the bends are fully extended and the local elastic response of A-tracts, is exposed. This force regime unveiled the extraordinary enthalpic rigidity of A-tracts.

pairs, which is energetically very costly. Conversely, the large stretch modulus found here can be regarded as indicative of a particular structure of the A-tracts that differs from the canonical B-DNA.

5.5.2 Biological implications

A-tracts appear enriched in the genomes of widely diverse organisms [164–167] and have been linked to several biological processes including recombination, replication and DNA packaging [119, 170]. In many of these processes, A-tracts are thought to operate by stabilizing the formation of DNA tertiary structures. For instance, A-tracts have been shown to localize supercoils [137] and might facilitate the formation of loops at regulatory regions [119]. In addition, short phased A-tracts might stabilize the highly bent structure adopted by DNA in the nucleosome [121, 196] and likely contribute to DNA packaging in the bacterial nucleoid [166]. However, somewhat surprising, in some other cases A-tracts seem to inhibit bent DNA conformations. Possibly the most prominent example are long (>10 bp) poly-A stretches, which are thought to resist strong bending and have been shown to significantly destabilize nucleosomes [122, 168].

Based on our results, we may propose an explanation for this apparent contradiction on A-tracts flexibility, with views on achieving a better understanding of the biological function of these sequences. We have shown that phased A-tracts greatly bend the DNA; however the tracts themselves appear rigid at high forces. This finding suggests that the relative magnitude of these two effects - intrinsic bending and A-tract rigidity – could be modulated by tuning the length and distribution of the A-tracts. Short ($\sim$ 5 bp) and phased A-tracts would amplify the intrinsic bending, whereas a long ($>$ 10 bp) individual A-tract would greatly stiffen the DNA. The former should then be involved in bent DNA conformations, such as in supercoils or nucleosomes; whereas the latter would be preferred for rigid sequences that preclude strong distortions in the DNA. This idea is in line with the findings reported in the literature [121, 122, 137, 145, 196], as described in the previous paragraph.

Furthermore, the considerations above raise the question of which forms of DNA flexibility are required for compaction into nucleosomes. In particular, if

the mechanical properties of poly-A stretches are the responsible for their nucleosome depleting character, one would expect a high bending stiffness of these sequences. On the contrary, what we found is that A-tracts are stiff in terms of stretching, but not bending. In addition, a recent study on the flexibility of poly-A stretches does not support the generalization that these sequences are always resistant to bend [183]. Therefore, we propose that stretching, which reflects a short-scale deformation [162], could play a more relevant role in nucleosome stabilization than the long-range WLC bendability. A similar idea has been proposed in simulation studies [62, 162]. The high positive charge of the histones could well induce tight bends in the DNA. However, the precise adaptation of the DNA molecule to the shape imposed by the histones relies on very particular local motions of the DNA [62, 197, 198]. A high stretch modulus might be indicative that these local motions are precluded. The hypothesis that bending is not determinant for nucleosome stabilization is supported by the structure of a nucleosome core particle containing a 16 bp-long poly-A. This tract is able to bend, but it adopts a distorted configuration that ultimately destabilizes the nucleosome [199]. Interestingly, CpG islands, another well-known example of nucleosome destabilizing sequence [123], have also been reported to show a large stretch modulus [134]. Finally, in addition to its role in nucleosome positioning, A-tract rigidity would mechanically stabilize the structural features of these sequences, which constitute a recognition mechanism in DNA-protein interactions [115, 121].

5.6 Conclusions

In this work we have investigated the mechanical properties of A-tracts at the single-molecule level using atomic force microscopy, magnetic tweezers and optical tweezers. Our AFM measurements evidence the intrinsic curvature of these sequences, as manifested by the deviation of their mechanical properties from the WLC model. We hence derived a variation of this model that accounts for polymers with a preferred curvature: the intrinsically-bent worm-like chain (IBWLC) model. This model described our AFM experimental data with exceptional accuracy, and allowed us to quantify the intrinsic and entropic bending

of these molecules. In addition, we employed MT and OT to characterize the response of A-tracts under the action of an external force. At low forces, the mechanical response of A-tracts is described by a persistence length of $\sim$ 20 nm, a hallmark of bent DNA. This reduction of the A-tracts persistence length was preserved for a wide range of NaCl concentrations, supporting that A-tracts bending does not rely on the interaction with ions. Interestingly, OT experiments revealed a $\sim$ 1.5-fold increase in the stretching modulus compared to random sequence DNAs. This exceptional stretching stiffness of A-tracts likely reflects the unusual local rigidity of these sequences which has been linked with their efficiency in nucleosome depletion. Our results reconcile contradictory views on A-tracts flexibility and stress the need of appropriate models and techniques that allow disentangling all the forms of DNA flexibility that may be involved across wide length scales. From a broader perspective, our comprehensive study on the mechanical properties of A-tracts might shed light on the intricate relation between sequence dependent flexibility and function in the double helix.

6 Sequence Induced Bending in dsRNA: AU-tracts

6.1 Introduction

In the last decades dsRNA has emerged as an important effector in processes such as the regulation of gene expression by micro RNAs (miRNA) [11], or the host responses to dsRNA encoded by viruses [12]. In these and other processes, proteins such as adenosine deaminases or Dicer-like enzymes bind specific dsRNA sequences but it is poorly understood how this recognition is achieved. Given the compact structure of the RNA duplex, with little exposure of the bases, it is conceivable that proteins recognize structural features of dsRNAs, such as helical deformations imposed by the sequence [160, 200]. In fact, some examples of proteins that recognize bent dsRNA owing to bulges or wobble base pairs have been reported [161]. From a more general perspective, sequence-dependent dsRNA conformations might also be relevant in the folding of RNA into complex tertiary and quaternary structures, such as riboswitches and ribosomal subunits.

In the case of double-stranded DNA (dsDNA) a plethora of works have enabled a comprehensive picture of how the nucleotide sequence affects the conformation and flexibility of the double-helix. Possibly the most representative example of such sequence-dependent deformations are the so-called A-tracts, runs of adenines and thymines that produce a bend in the DNA [26, 70, 119, 170, 175]. These sequences have been largely discussed in Chapter 5. In short, when several A-tracts occur in phase with the helical pitch, their combined effects yield a significant global curvature of the dsDNA [26]. At the molecular level, A-

tracts show a peculiar conformation distinct from B-DNA, with a characteristic narrow minor groove [25].

In contrast to the exhaustive literature on sequence-dependent dsDNA bending, much less is known about how the sequence affects dsRNA conformations. Early crystallographic works reported helical kinks in the structure of an RNA duplex consisting of alternating adenines and uracils [201, 202]. However this bent conformation was stabilized by the intermolecular interactions among the molecules forming the crystal and, therefore, bending could not be attributed to the RNA duplex alone. Later on, two independent works studied the effect of phased A-tracts on dsRNA bending and showed that, contrary to dsDNA, A-tracts do not induce a global curvature in dsRNA [203, 204]. As a result of these works, it is frequently accepted that dsRNA is a uniform helix with little sequence variation.

Here we present a combined computational and experimental approach to address the problem of sequence-dependent bending in dsRNA. By means of molecular dynamics simulations, we performed a systematic analysis of how the sequence affects the structure of dsRNA. Our simulations predicted that a sequence motif, that we named AU-tracts, cause a bend in the RNA double-helix. This prediction was experimentally tested by atomic force microscopy imaging, which showed that AU-tracts promote bent conformations in dsRNA. Our work unveils that the bendability of dsRNA is sequence dependent, challenging the traditional picture of dsRNA as a regular helix.

6.2 The dsRNA sequence affects the size of the major groove

In order to explore sequence effects in dsRNA, we analyzed the unrestrained benchmark sequences simulations from Chapter 4 (see Table 6.1). We first measured the size of the grooves of these sequences. The grooves are the regions where interactions with proteins usually take place, and in the case of dsDNA their size has been established as a critical factor in specific recognition [12]. Using the software Curves+ [86], we found that the major groove width is highly dependent on the sequence, being able to change by as much as 7 Å. This is shown in Fig. 6.1a, where we represent the values of the major groove

Benchmark Sequences		
Label	Sequence	Bending (deg)
Poly-A	$G_4(AA)_8G_4$	2.1
Poly-AC	$G_4(AC)_8G_4$	2.2
Poly-AG	$G_4(AG)_8G_4$	1.6
Poly-AU	$G_4(AU)_8G_4$	3.6
Poly-CG	$G_4(CG)_8G_4$	3.6
Poly-G	$G_4(GG)_8G_4$	1.8
AU-tract Length		
Label	Sequence	Bending (deg)
AU-3	$G_{11}AUAG_{11}$	7.9
AU-4	$G_{10}(AU)_2G_{10}$	12.2
AU-5	$G_{10}(AU)_2AG_{10}$	12.8
AU-6	$G_9(AU)_3G_9$	14.4
AU-7	$G_9(AU)_3AG_9$	14.2
Testing Sequences		
Label	Sequence	Bending (deg)
Seq. 1	CUAGAUGAGAGAUUCGGCUGUCAG	4.6
Seq. 2	CAGAGCUUAGCUGAUUGGUGAACC	0.9
Seq. 3	GCUGGUUUCCGCGGGUGGUUUAGA	2.0
Seq. 4	GCUGGUUUCAUAGGGUGGUUUAGA	10.0
Seq. 5	UUUAUUGGUGGUUUAUAAUGUGCG	11.9
Seq. 6	GCUGGUUUCAUAUGGUGGUUUAGA	15.9

Table 6.1: DsRNA sequences studied for benchmarking and testing AU-tract bend. DsRNA sequences are represented in abbreviated form without the complementary strand and written from the 5' end to the 3' end. The benchmark sequences are the same as in Chapter 6. AU-tract length sequences are all of approximately the same size (24 or 25 bp depending on whether the AU-tract comprised an even or odd number of bp) and contain an AU-tract exactly in the center of the sequence. These centered AU-tracts were of varying lengths from three (AU-3) to seven (AU-7) base pairs. Testing sequences were designed to contain other potential bending motifs, such as AUU, UAA or CGCG (highlighted in green, Seqs. 1-3) and to include AU-tracts in different contexts (highlighted in red, Seqs. 4-6).

width along the helical axis for the benchmark molecules. Notice that, since these sequences consist of repeating dinucleotides, their major groove should be regular along the helical axis, as manifested by the flat lines of Fig. 6.1a. This constitutes a convergence test for our simulations.

Notably, we observed that the major groove is primarily responsible for modulating the main helical parameters of the molecules. Specifically, the helical rise and helical twist are highly correlated (R=0.987) and anticorrelated (R=-0.974), respectively, with the major groove width (Fig.6.1b, c). These results show that the primary sequence can simultaneously induce an elongation and unwinding of the dsRNA by expanding the major groove. Conversely, contraction of the major groove results in shrinkage and overwinding of the double helix. This mechanism is illustrated in Fig. 6.1d, where we computed the average structures over the 1 μs simulation time of the two sequences with extreme values of major groove width: the poly-AU and the poly-CG. The former is the most compact sequence with a very narrow major groove. As the major groove is enlarged, the molecule approaches a stretched and unwound conformation, which is maximal in the poly-CG sequence. The helical rise of the poly-CG and the poly-AU molecule was $\pm$ 10 $\%$ of the canonical value of the extension per base pair of dsRNA, which is $\sim$2.9 Å [17]. Importantly, sequence variations in dsRNA compactness have been found to be consistent for different force field and water model choices [74, 75].

6.3 AU-tracts bend the RNA duplex by compressing the major groove

We have shown that the sequence of homogeneous dsRNA molecules modulates the major groove size, with the poly-AU sequence leading to a significant compaction of the major groove width. Compaction of the minor groove in DNA A-tracts has been linked with the intrinsic bending induced by these sequences [119]. Motivated by the DNA case, we next explored if major groove narrowing can lead to bent structures in dsRNA. In order to do so, the poly-G sequence, which shows standard values of the structural parameters (Fig. 6.1), will be modified to include a stretch of alternating A's and U's, hereafter AU-tract.

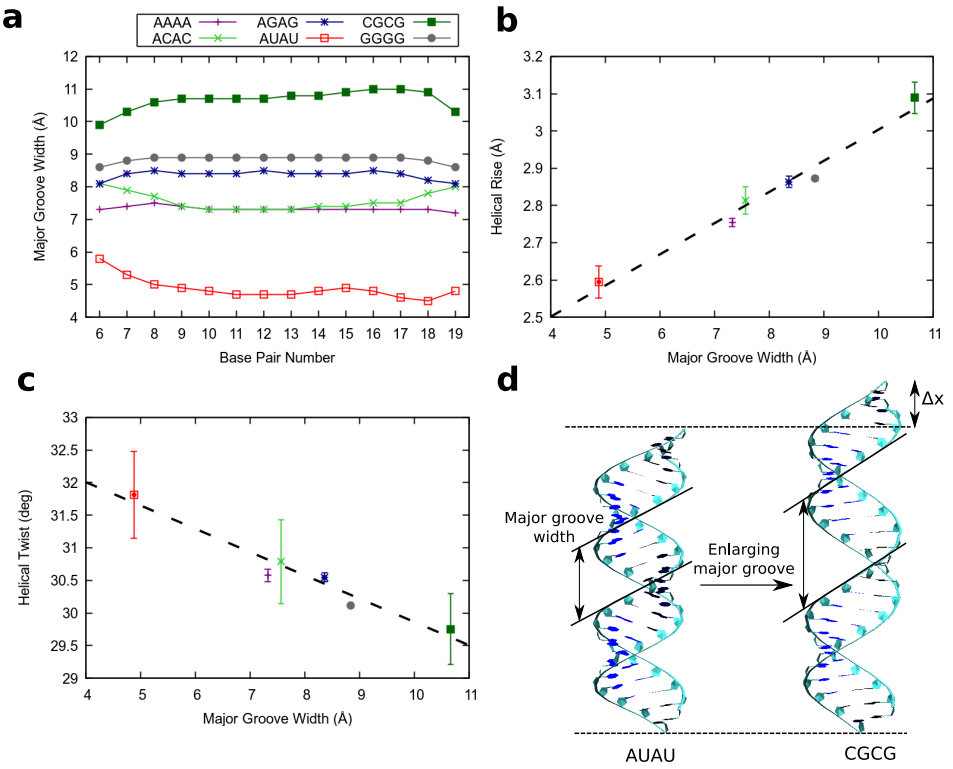

Figure 6.1: The major groove width tailors the overall structure of the dsRNA helix. a, Value of the major groove width measured along the helix for the benchmark sequences. The average structures over the simulation time were computed for the benchmark molecules using the software AmberTools [76]. These structures were then analyzed with the software Curves+ [86] to obtain the values of the major groove width along the sequence. b, Helical rise and c, helical twist of the benchmark molecules measured as a function of the major groove width. Values of the helical rise and helical twist were obtained for each base pair step from the average structures of the benchmark molecules using the software Curves+ [86]. Then these values, together with the major groove width values from panel a, were averaged over the 15 central base pair steps. The legend is the same as in panel a. Error bars are the standard error of the mean. X-axis error bars are within the symbols. The dotted line represents a fit of the data to a linear function. d, Average structures of the poly-AU and poly-CG over the simulation time. These structures illustrate how the sequence induces an elongation in the molecule by enlarging the major groove.

We will consider five different sequences with AU-tracts of lengths varying from three to seven nucleotides, that will correspondingly be denoted as AU-3 to AU-7 (Table 6.1). All sequences were designed to be similar in length (24 or 25 base pairs) and to contain the AU-tract located exactly in the center of the duplex.

Our results revealed a localized compression of the major groove at the position of the AU-tract. This can be seen in Fig. 6.2, where we represented the major groove width profiles of the poly-G (same as Fig. 6.1a) and the AU-tract containing sequences. The homogeneous major groove width of the poly-G molecule contrasts with the abrupt drop found when an AU-tract is introduced. As the length of the AU-tract increases, the reduction of the major groove width is more pronounced, reaching lower values and extending over longer distances along the duplex. Interestingly, a short 3bp-long AU-tract is enough to induce major groove narrowing, although the effect is significantly smaller than for longer AU-tracts.

Strikingly, compression of the major groove by AU-tracts results in bent dsRNA structures. This can be noticed by visual inspection of the computed average structures of the molecules throughout the simulation time. The average structures of the AU-tract containing molecules presented a bend at the position of the tract and were therefore unable to be embedded inside a virtual cylinder. On the contrary, the poly-G sequence was straight and, therefore, could be fitted inside a cylinder (see Fig 6.2). In order to provide a quantitative description of this bending we resorted to the curvilinear axis definition of Curves+ [86]. The bending angle was computed from the average structures of the AU-tracts sequences, discarding the four base pairs adjacent to each molecule's termini. This bending angle is plotted as a function of the AU-tract length in Fig. 6.3, where we also included the poly-G homopolymer, which can be regarded as a zero-length AU-tract. These measurements corroborated the bending effect of AU-tracts that we inferred from visual inspection of the dsRNA structures. The poly-G sequence, which lacks AU-tracts, was found to be essentially straight, as quantified by a very small bending angle of $\sim$2 deg. Interestingly, the shortest AU-tract considered, which is only 3 bp long, already induced a significant bending of $\sim$ 8 deg in the RNA duplex. This bending increases with the AU-tract

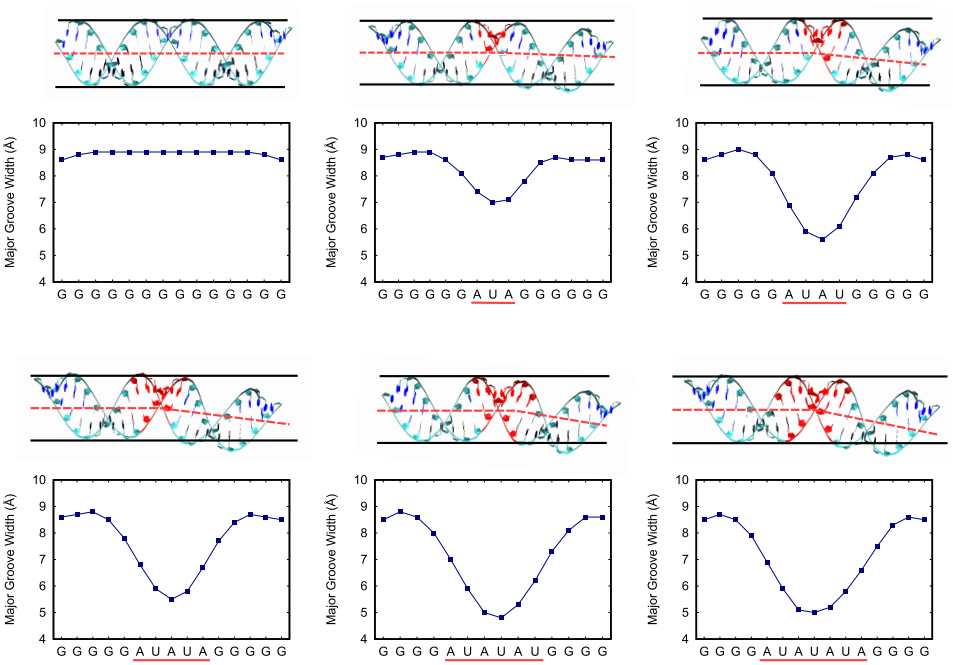

Figure 6.2: Structure and major groove of sequences containing AU-tracts of different lengths. Structures of poly-G sequence and AU-3 to AU-7 are represented with the AU-tract highlighted in red. The black lines represent a cylinder, which is unable to embed highly bent molecules, namely those containing an AU-tract. An approximate helical axis was drawn in red dotted line to guide the eye. Major groove width profiles of poly-G and AU-3 to AU-7 were computed and represented as in Fig. 6.1a. Localized drops in these profiles are found in the AU-tracts (underscored in red) which coincide with the bending region of the molecule (top).

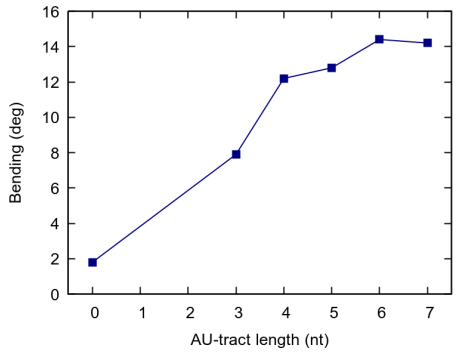

Figure 6.3: AU-tracts bend a homogeneous dsRNA sequence. Bending was computed for the AU-tracts of different lengths using the curvilinear helical axis from the software Curves+ [86] and neglecting four base pairs on each terminal of the molecule. A line connecting the points was drawn to guide the eye.

length, reaching a maximum of 14.4 deg at 6 bp and then slightly decreasing at 7 bp. However, we may assert that for AU-tracts of four base pairs or longer, the bending angle is roughly constant and variations are within the error of the measurement.

6.4 AU-tracts induce bending in random dsRNA sequences

The sequences studied so far have allowed us to unveil the phenomenon of AU-tract bending and to relate the length of the tract with the magnitude of the bending. Nevertheless, these model sequences – repeating dinucleotides or mononucleotides – do not reflect realistic dsRNA molecules. We therefore designed a new set of simulations aimed at exploring the role of AU-tract bending in arbitrary dsRNA sequences as well as testing other potential bending motifs. All these testing sequences comprise 24 bp, see Table 6.1. As in previous sections, simulations were extended to 1 μs time and bending was evaluated from the average structures discarding the four terminal base pairs from each side.

Remarkably, among the studied sequences, only those containing AU-tracts were bent. This is shown in Fig. 6.4, together with the major groove width profiles of the testing sequences. Seq. 1 and Seq. 2 contain no AU-tract, but

other motifs rich in A's and U's, namely AUU and UAA. Contrary to AU-tracts, these motifs produced very modest variations in the major groove width and, consequently, Seq. 1 and Seq. 2 were nearly straight (Fig. 6.4). Seq. 3 presented a CG-tract that locally enlarged the major groove, in line with the results from Fig. 6.1. However, this effect was not translated into an enhanced bending of the duplex. As anticipated, only sequences with AU-tracts, namely Seqs. 4-6, were significantly bent, all showing values of the bending angle larger than 10 deg (Fig. 6.4b). Among these three sequences, Seq. 4, which has the shortest AU-tract – 3 bp long – scored the lowest bending angle, but still substantially larger than any of the sequences lacking AU-tracts (Seqs. 1-3). Consistently, Seq. 4 presented a less pronounced drop in the major groove width compared with Seq. 5 and Seq. 6, which contain a longer AU-tract comprising four base pairs.

The results here exposed are in line with crystallographic studies reporting a bend in duplexes containing a central AU-tract [201, 202, 205]. However, crystal packing can induce spurious bending in nucleic acids [177] and, indeed, the bent helices observed in the AU-tract structures were partly attributed to intermolecular interactions among different duplexes of the crystal [201, 202]. In the following section we will provide experimental evidence that AU-tracts promote the formation of bent dsRNA structures at the single-molecule level.

6.5 Experimental evidence of AU-tract induced bending

Motivated by our simulation findings, we performed atomic force microscopy (AFM) imaging to experimentally test the effect of AU-tracts on dsRNA bending. We speculated that AU-tracts located in phase with the dsRNA helical pitch would amplify their bending, similar to the case of A-tracts in the DNA sequences (see Chapter 5). Therefore, we fabricated a 612 bp dsRNA construct that contains phased AU-tracts with a periodicity of 11 bp. As control, we considered an arbitrary dsRNA sequence of the same length and with a GC-content of $\sim$50%. DsRNA molecules were fabricated by Dr. Clara Aicart-Ramos from Fernando Moreno-Herrero's lab.

Using the experimental conditions presented in Section 1.5.2, we obtained

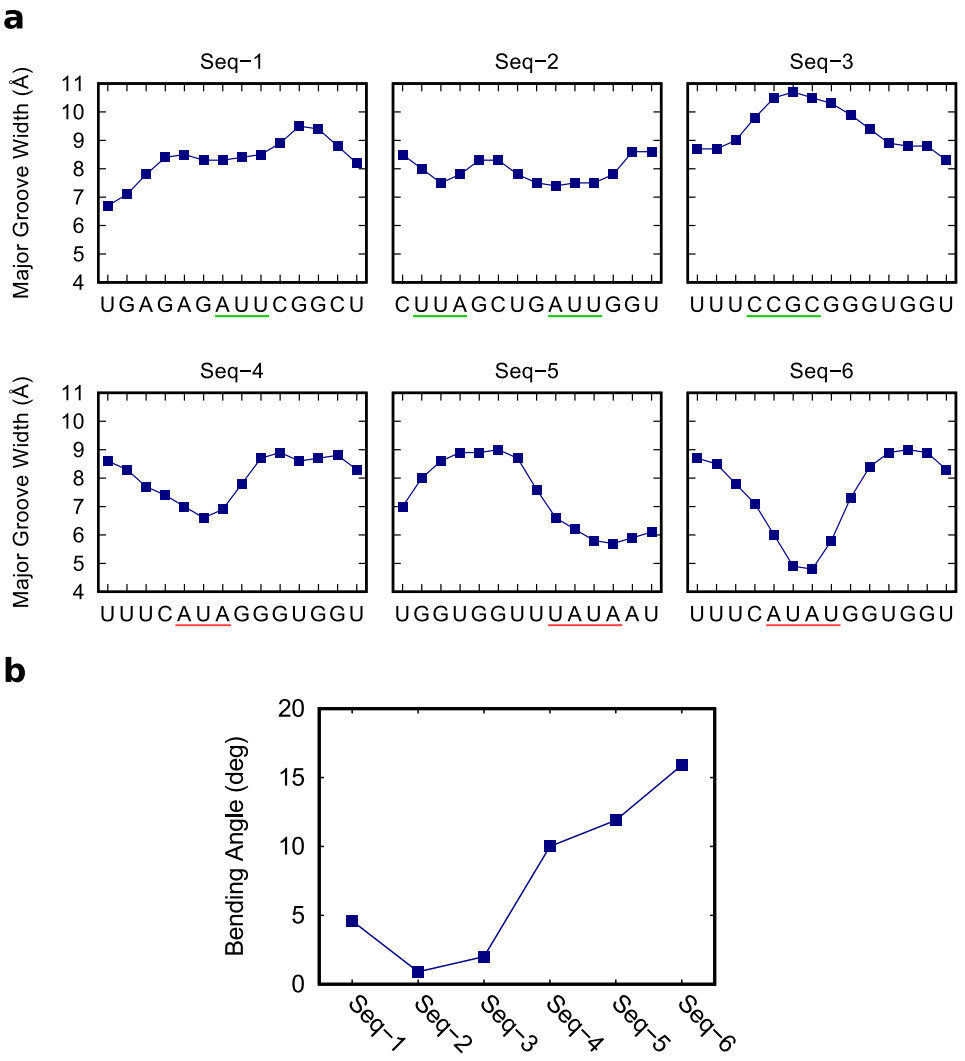

Figure 6.4: AU-tracts bend arbitrary dsRNA sequences. a, Major groove width profiles of the arbitrary sequences were computed and represented as done in Fig. 6.1a. AU-tracts were highlighted in red and other tested motifs, such as AUU or CGCG, were highlighted in green. b, The bending angle was measured for the arbitrary sequences as done in Fig. 6.3. Only Seq 4-6, which contain an AU-tract, scored a bending angle larger than 10 deg.

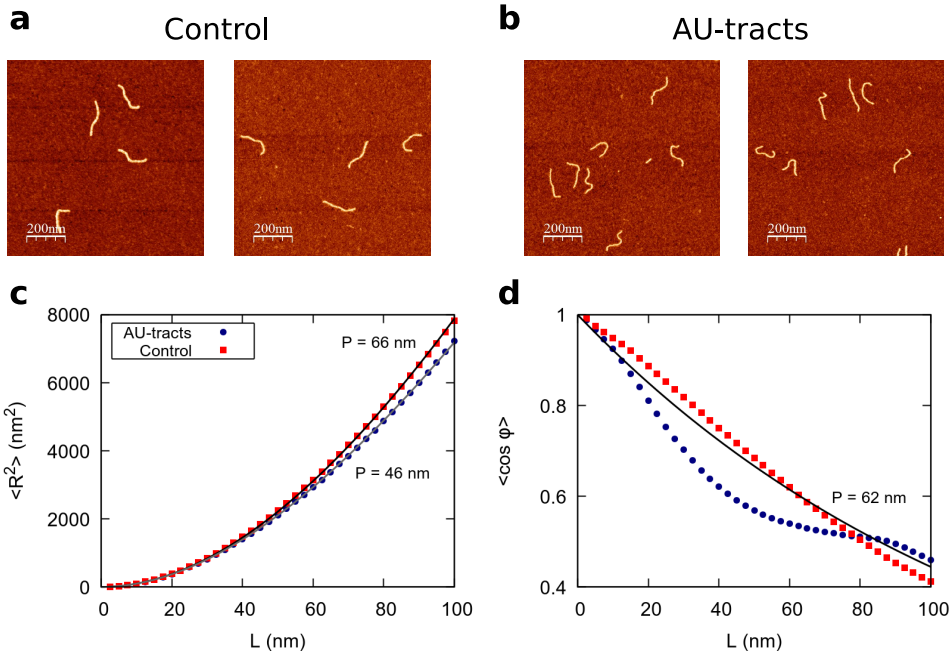

Figure 6.5: Experimental evidence of AU-tract bending. Representative AFM images of control (a) and AU-tracts (b) dsRNA molecules. (c) $< R^2_{s,s+L} >$ was plotted as a function of the contour distance, L, between two points. The black and the gray lines represents the fits to Eq. 6.1 of the control (red) and AU-tracts (blue) data, respectively. (d) $< \cos\varphi >$ as a function of L. The control data was fitted to Eq. 6.2.

AFM images of control and AU-tracts dsRNA molecules, see Fig. 6.5a, b. From the AFM images, we measured the contour lengths of the molecules and obtained values of 179 $\pm$ 3 nm and 179 $\pm$ 4 nm for the control and AU-tracts respectively. These values yield a ratio of 2.9 Å/bp, which coincides with crystallographic data of dsRNA [17] and with our MD simulations (Fig. 6.1b).

A thorough analysis of our AFM images revealed that AU-tracts molecules are indeed more bent than control dsRNA molecules of arbitrary sequence. In order to quantitatively assess the bendability of the control and AU-tracts dsRNA molecules, we followed the protocol presented in Section 1.5.2 and used in Section 5.2.1 for the characterization of DNA A-tracts. We first obtained traces of points separated by l= 2.5 nm that follow the trajectory of the molecules [30, 89]. From these traces, we computed the mean squared end-to-end distance,

$< R_{s,s+L} >$, and the mean tangents' cosine $< \cos\varphi >$ for points separated by a contour distance L. According to the WLC model

$$< R^2_{s,s+L} > = 4P\left(L + 2P\left(e^{-L/2P} - 1\right)\right) \tag{6.1}$$

$$< \cos\varphi > = e^{-L/2P} \tag{6.2}$$

where P is the persistence length and is directly proportional to the bending rigidity of the polymer $P = B/k_BT$.

We firstly fitted the $< R^2 >$ data of the control dsRNA to Eq. 6.1 and obtained a persistence length of $P = 66 \pm 1$ nm, consistent with previous single-molecule experiments on arbitrary dsRNA sequences [32, 37, 99], see Fig. 6.5c. However, the AU-tracts molecule presented a persistence length of $P = 46 \pm 1$ nm, around 30% lower than that of the control. This finding indicates that dsRNA molecules with phased AU-tracts are more prone to bending deformations than arbitrary sequences, supporting our simulation results. Furthermore, analysis of the tangents correlations exposed deviations of the AU-tracts molecules from the standard WLC model, see Fig. 6.5d. The control data could be fitted to the WLC model (Eq. 6.2) and yielded a persistence length of $P = 62 \pm 1$ nm. Although slightly lower, this value of P is consistent with the one obtained by fitting the $< R^2 >$ data. On the other hand, the AU-tracts data did not fit to the WLC model, indicating that the AU-tracts molecules present more complex mechanical properties than arbitrary dsRNA sequences. Further attempts to fit the AU-tracts data to the IBWLC developed in Section 5.2.2, were unsuccessful. Future work is required to fully experimentally characterize the mechanical properties of these intriguing sequences.

6.6 Discussion

6.6.1 AU-tracts: similarities and differences with DNA A-tracts

As discussed in Chapter 5, sequence-dependent DNA bending takes place by means of A-tracts: sequences of at least four A·T base pairs without a TA step, which display a particular conformation that differs from that of canonical B-DNA [25, 119]. Here we compared the structural features of the dsRNA AU-tract

of Seq-6 with the DNA A-tract from a recent high-quality NMR structure of the Drew-Dickerson Dodecamer (DDD) [206]. The DDD is the most extensively characterized DNA duplex and contains a central A-tract: CGCGAATTCGCG. The analysis of the most relevant structural parameters is presented in Fig. 6.6 and revealed intriguing similarities and differences between A-tract and AU-tract bending. The central region of the DDD shows the standard features of A-tracts, which are a highly negative propeller twist, a narrow minor groove and a negative roll [207–209]. Moreover, the major groove width showed little variation and all the plots were symmetric, a consequence of the palindromic sequence of the DDD. Similar to the A-tract case, a large negative propeller twist was observed in the AU-tract. However, the roll parameter presents very different trends in the two molecules. This parameter presents a maximum at the center of the AU-tract, but a minimum in the A-tract. This difference in roll can be linked to the changes observed in the dimensions of the grooves. Positive values of roll are attributed to bending towards the major groove and, therefore the increase in roll is consistent with the compression of the major groove observed in the AU-tracts (see Fig. 6.2 and Fig. 6.4). Conversely, the highly negative roll found in A-tracts can be associated with their minor groove narrowing, a well-known feature of these sequences [207–209]. Furthermore, notice that only one of the grooves showed a significant sequence variation. Namely, the minor groove of the AU-tract and the major groove of the A-tract were approximately homogeneous.

In addition, A-tract bending has traditionally been characterized by anomalous gel migration. Namely, DNA molecules containing phased A-tracts migrate slowly in electrophoretic gels when compared to arbitrary DNA molecules of the same length. We checked if AU-tracts also caused anomalous migration of dsRNA in agarose gels, but we found no difference when comparing the AU-tract molecule with the control, see Fig. 6.7. It is therefore likely that the effect of AU-tract bending in dsRNA is less pronounced than that of A-tract in DNA. This hypothesis is supported by our AFM measurements. We found that the AU-tract molecule presents a lower persistence length – and thus a higher bendability – than the control dsRNA. However, this reduction was of $\sim$30%, much smaller than the reported decrease of $\sim$50% when comparing the estimated persistence

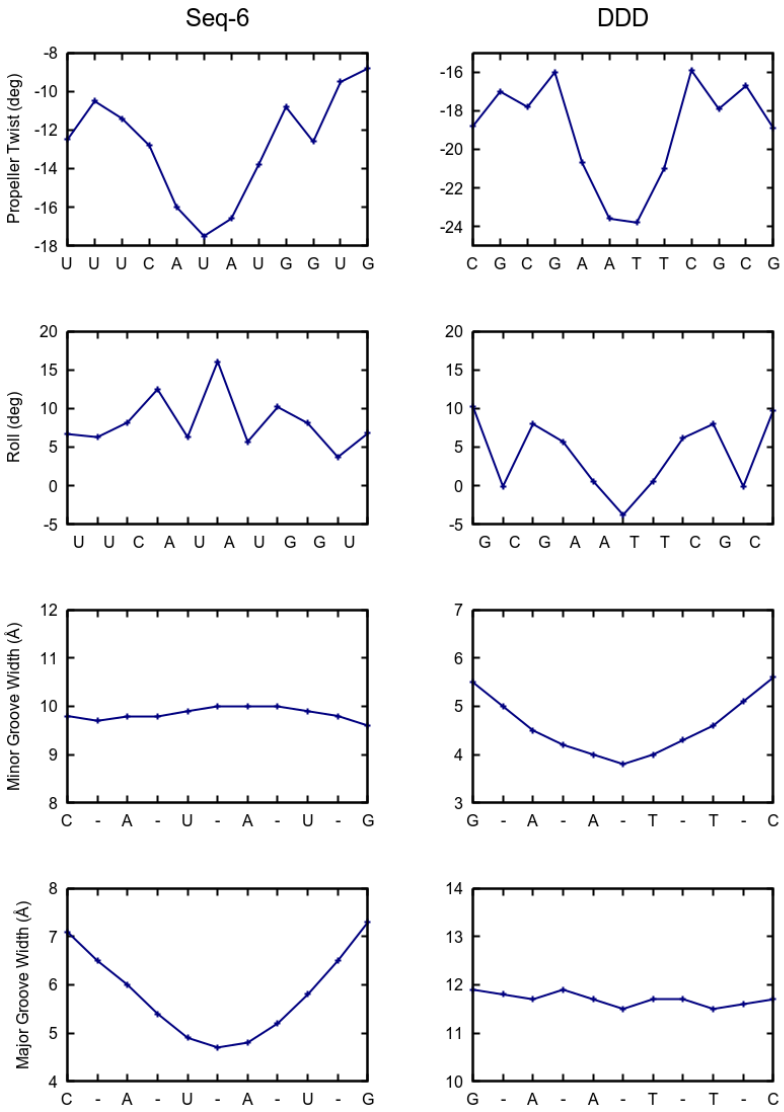

Figure 6.6: Comparison of structural parameters of dsRNA AU-tracts and dsDNA A-tracts. The average structure of the simulated Seq 6 (Table 6.1) and a recent NMR [206] structure of the Drew Dickerson Dodecamer (DDD) were taken as representative examples of a dsRNA AU-tract and a dsDNA A-tract respectively. The analysis was performed using Curves+ [86]. In addition to the major groove width, the structural parameters propeller twist, roll and minor groove width, which are typically used in the characterization of A-tracts [203] were computed.

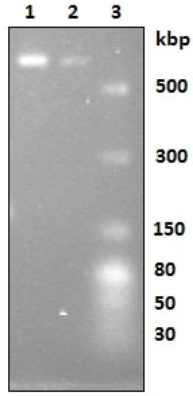

Figure 6.7: Migration of control and AUAU-tracts dsRNA substrates in agarose gels. 2% agarose gel electrophoresis of 612 bp-dsRNA AFM substrates (AU-tract, lane 1, and control DNA, lane 2) were run (40 ng/lane) in 1xTBE buffer. Lane 3 corresponds to the dsRNA ladder. The electrophoresis was run at room temperature for 100 min at 60 V (4V/cm) in the absence of intercalator and the gel was later stained with SYBR®Safe. Both substrates show the same migration.

length of phased A-tracts with an arbitrary DNA sequence (Section 5.2.1).

6.6.2 Implications for dsRNA recognition

DsRNA structures are ubiquitous in cells and, together with dsRNA-binding proteins, are central players in cellular processes, such as mRNA biogenesis and editing, microRNA processing and function as well as anti-viral defense [210–213]. Different classes of dsRNA-binding proteins exist, and some have a highly conserved 60-70 amino acid $\alpha\beta\beta\beta\alpha$-fold dsRNA-binding motif (dsRBM, also referred to as dsRBD) [160, 210]. In general, the dsRBM-containing dsRNA binding proteins recognize the A-form of dsRNA through the 2'-OH groups of riboses and the width of the minor and major grooves. The binding of the dsRBM to dsRNA has traditionally been considered as sequence-independent, however recent studies have suggested a possibility for a limited dsRNA sequence specificity, although the recognition mechanism still poorly completely understood

[160]. In this context, sequence-dependent dsRNA conformations are likely to play an important role. For example, a selective dsRNA sequence triggers the binding of the cytoplasmic DExD/H-box helicase dsRNA sensor RIG-I, where two G-U wobble base pairs in a short 8 bp-long hairpin functionally activated RIG-I by giving rise to a bent helix [214]. Two other hairpin sequences with interhelical bends were also found to activate RIG-I. Bends or bulges in dsRNA molecules have further been linked to recognition by the antiviral agent PKR [161], or have been found in structures of dsRNA-protein complexes [215–217]. One can thus speculate that the dsRNA curvature at the AU-tracts observed in the present study might contribute to specific target recognition by dsRNA-binding proteins. This would add to other mechanisms of dsRNA sequence recognition, such as the recently proposed contacts through the minor groove in case of dsRBD containing proteins [218].

6.7 Conclusions

In this work we have presented a rationale of sequence effects in bending the RNA double-helix. Our benchmark sequences simulations revealed that the poly-AU sequence adopts a characteristic conformation with a narrow major groove. By inserting AU-tracts of different lengths inside a poly-G homopolymer, we found that a 3 bp-long AU-tract is enough to induce a bend, but the optimal length is four base pairs. This finding was consistent when studying AU-tracts located in arbitrary sequences. Finally, our simulation results guided the fabrication of dsRNA constructs suitable for measuring the effect of AU-tract bending in single-molecule experiments. Using AFM imaging, we found that these AU-tract molecules are more bendable than control dsRNA molecules of arbitrary sequence, supporting the prediction of our simulations.

The results exposed in this work allow us to propose a simple model for AU-tract bending, which is schematically illustrated in Fig. 6.8. Molecules with a homogeneous major groove, such as the poly-AU and poly-CG, will show no or very little bending, and modulation of the groove width will only result in elongation and untwisting of the molecule. On the contrary, molecules containing an AU-tract can be highly bent, as a consequence of their irregular major

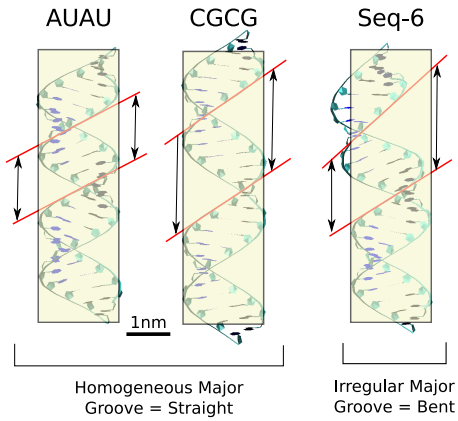

Figure 6.8: Structural model proposed for intrinsic bending in dsRNA. The molecules shown are the computed average structures of molecules poly-AU, poly-CG and Seq-6. A small/large homogeneous major groove width (poly-AU and poly-CG sequences) will only result in shrinking/elongation of the molecule, but no bending. On the contrary, localized compression of the major groove, as observed in AU-tracts, are proposed to induce a bend in the dsRNA helix.

groove distribution. Due to its simplicity, we believe that this model is a valuable starting point towards elucidating many unknown features of AU-tracts. For instance, it remains to be explored the role of ions, water molecules or potential hydrogen bonds in stabilizing the AU-tract. In addition, future experimental works should determine whether, similar to DNA A-tracts, AU-tracts are able to induce a static bend in the double-helix.

Intrinsic bending induced by dsDNA A-tracts has been linked to multiple biological functions such as nucleosome positioning, localization of supercoils or germ-line gene silencing. It is therefore expected that the sequence-dependent bending reported here for dsRNA might also have important biological implications. Most notably, AU-tracts might provide a mechanism for sequence recognition based on dsRNA shape. Hence, our results might shed new light on how dsRNA sequence dictate specificity on dsRNA-protein interactions, and thereby impact on biological processes ranging from antiviral response to gene silencing.

7 Conclusions and Future Perspectives

7.1 General Conclusions

In this PhD dissertation we have performed molecular dynamics (MD) simulations and single-molecule experiments to explore the mechanical properties of double-stranded DNA (dsDNA) and RNA (dsRNA). The powerful combination of these two techniques enabled us to unravel complex molecular mechanisms that underlie the dynamics of the DNA and RNA duplexes. The conclusions of this dissertation are

1. We developed a protocol to implement a constant force during an MD simulation. We used these constant-force MD (CFMD) simulations to determine the mechanical parameters of a dsDNA and dsRNA molecule of arbitrary sequence, in good agreement with previous experiments. In particular, we reproduced the opposite twist-stretch coupling of these two duplexes. We then explained this phenomenon at the atomic level: upon stretching, the dsDNA inter-strand distance decreases, allowing the molecule to overwind. However, this is not the case for dsRNA, for which the inter-strand distance remains unchanged as the molecule elongates, resulting in unwinding. The capability of dsDNA to overwind when stretched was linked to the larger flexibility of its sugar group.

2. We employed CFMD simulations to study the structure and the stretching flexibility of several dsDNA sequences. We observed that the dsDNA extension is dependent on the sequence, as a consequence of a curvature of the molecule that we named *crookedness*. The *crookedness* was closely

related with the stretching flexibility of the duplex: dsDNA molecules with a higher crookedness curvature (and thus shorter extension) presented a softer stretching response, and vice versa. This one-to-one correspondence has been rationalized using a simple phenomenological model.

3. We studied the sequence-dependent mechanical response of dsRNA using CFMD simulations. We found that the stretching flexibility of dsRNA is highly sequence-dependent. However, unlike the dsDNA case, these sequence variations were not governed by the *crookedness* curvature. Instead, the pyrimidine-purine steps, which are highly flexible, dominated the elastic response of arbitrary dsRNA sequences. We proposed that the sequence regulates in a completely different manner the flexibility of dsDNA and dsRNA duplexes, such that sequences that are highly flexible in the dsDNA case can be rigid in dsRNA and vice versa.

4. Using a combination of single-molecule techniques, we characterized the mechanical properties of a sequence from the *C. elegans* genome that contains phased A-tracts. Atomic force microscopy (AFM) imaging of these molecules revealed a long-range bending that extends to distances larger than 200 nm. We explained this phenomenon by means of a polymer model that accounts for the intrinsically bent structure of the individual A-tracts. Magnetic tweezers force spectroscopy experiments showed that, at forces lower than 10 pN, these A-tracts molecules presented a peculiar mechanical response, which differed from standard entropic models. Finally, we unveiled the large stretch modulus of A-tracts by means of optical tweezers.

5. We assessed the effect of the nucleotide sequence on dsRNA bending. Our MD simulations predicted that sequences consisting of alternating adenines and uracils, that we named AU-tracts, would compress the mayor groove of the RNA duplex, resulting in a bent structure. These AU-tracts needed to be at least three base pairs long to induce a bend in the dsRNA molecule, but the optimal length was four base pairs. We then sought an experimental confirmation for this effect. Using an AFM, we imaged

dsRNA molecules containing periodic repetitions of an AU-tract. We observed that these molecules were more bendable than dsRNA molecules of arbitrary sequence. This observation supported the computational predictions.

7.2 Future Perspectives

The results presented in this Thesis motivate further inquiry in certain topics. In the following, we expose the most interesting points that could be addressed in the near future.

- In Chapter 3 we showed that hypermethylation of a poly-CG sequence increases its stretching flexibility. However, previous works claimed that methylation of dsDNA CpG steps should make the duplex stiffer in the context of cyclization and nucleosome formation [142]. Two possible explanations exist for this conundrum. The first is that cytosine methylation affects bending and stretching deformations in different ways. Thus, methylation might favor stretching deformations, but make dsDNA rigid to bending. The other - more appealing - solution is that methylation alters the flexibility of dsDNA in a sequence-dependent manner. In this case, the sequence context of the CpG step and the distribution of the methylated cytosines (*e.g.* hemimethylation versus hypermethylation) would have affect the impact of methylation on dsDNA flexibility. Future works should clarify this issue.

- In Chapter 5, the force response of the A-tracts molecules in the $F < 10$ pN regime showed clear deviations from the worm-like chain (WLC) model. So far, we have not found a theoretical model that fully describes our experimental data. In the simpler scenario where no external force is present, finding an analytical solution to the IBWLC was already highly non-trivial. Therefore, it seems unlikely that an analytical formula can be found for an IBWLC under the action of an external force. Coarse-grain simulations might be the most suitable technique to approach this problem.

- In Chapter 6, we studied how AU-tracts induce bending in dsRNA. Our MD simulations and AFM experiments, as well as previous X-ray crystallography works, agreed that AU-tracts bend the RNA duplex. However, this bending was not observed in the electrophoretic gels. A plausible reason for this apparent discrepancy is that the AU-tract bend is only present above a threshold concentration of cations, which is larger than the concentration used in the gels. This was indeed the case of the G_3C_3 dsDNA sequence [219]. A crystal structure showed that this sequence bends the dsDNA [220], but this effect could only be detected in gels when these were run at high enough concentrations of divalent ions [219]. Such strategy could be followed to test the dsRNA AU-tract bending in gels.

Conclusiones y Perspectivas Futuras

7.3 Conclusiones Generales

En esta Tesis doctoral hemos estudiado las propiedades mecánicas del DNA y RNA de doble hebra (dsDNA y dsRNA, por sus siglas en inglés) mediante simulaciones de dinámica molecular (MD) y experimentos de molécula individual. La combinación de estas dos técnicas nos ha permitido desvelar complejos mecanismos moleculares que rigen la dinámica del dsDNA y del dsRNA. Las conclusiones de esta Tesis doctoral son:

1. Desarrollamos un protocolo que nos permitía introducir una fuerza constante durante el transcurso de una simulación de MD. Usando estas simulaciones de MD a fuerza constante (CFMD, por sus siglas en inglés), medimos los parámetros mecánicos de moléculas de dsDNA y dsRNA de secuencia arbitraria, obteniendo un buen acuerdo con experimentos previos. En concreto, logramos reproducir la respuesta torsional opuesta del dsDNA y el dsRNA ante una fuerza externa. Además, explicamos este fenómeno a nivel atómico: al ser estirado, la distancia entre las hebras del dsDNA se redujo, lo que permitió a la molécula enrollarse. Sin embargo, este no fue el caso del dsRNA, el cual, al ser estirado mantuvo la distancia entre las hebras constante y se desenrolló. Finalmente, relacionamos la capacidad del dsDNA de enrollarse al ser estirado con la mayor flexibilidad de su grupo azúcar en comparación con el dsRNA.

2. Usamos simulaciones de CFMD para estudiar la estructura y la flexibilidad de estiramiento de varias secuencias de dsDNA. Observamos que la extensión de las moléculas de dsDNA era dependiente de la secuencia,

debido a una curvatura de la molécula que denominamos *crookedness*. Este *crookedness* estaba estrechamente relacionado con la flexibilidad de estiramiento de la doble hélice: moléculas de dsDNA con una curvatura de *crookedness* mayor (es decir, una extensión menor) presentaron una mayor flexibilidad de estiramiento, y viceversa. Desarrollamos un modelo fenomenológico sencillo para explicar este efecto.

3. Estudiamos cómo la secuencia afectaba la respuesta mecánica del dsRNA usando simulaciones de CFMD. Encontramos que la flexibilidad de estiramiento del dsRNA era altamente dependiente de la secuencia. Sin embargo, al contrario del dsDNA, esta dependencia no venía dictada por la curvatura de *crookedness*, sino por los dinucleótidos pirimidina-purina, que son muy flexibles. En base a esto, comprobamos que la secuencia regulaba de maneras completamente diferentes las dinámicas del dsDNA y del dsRNA. Por ejemplo, una secuencia podía ser muy flexible en el caso del dsDNA pero rígida en el dsRNA, y viceversa.

4. Combinamos varias técnicas de molécula individual para caracterizar las propiedades mecánicas de una secuencia del genoma de *C. elegans* que contiene repeticiones de *A-tracts* en fase. Nuestros experimentos de microscopía de fuerza atómica (AFM) revelaron que estas secuencias presentaban un doblamiento que se extiendía a largas distancias de hasta 200 nm. Explicamos este fenómeno mediante un modelo de física de polímeros que tenía en cuenta la estructura doblada de los *A-tracts*. Los experimentos de pinzas magnéticas mostraron que, para fuerzas menores de 10 pN, los *A-tracts* poseían una respuesta mecánica peculiar que, a diferencia de moléculas arbitrarias de dsDNA, no podría ser explicada por modelos entrópicos. Finalmente, exploramos el régimen entálpico de los *A-tracts* mediante experimentos de pinzas ópticas y descubrimos que estas secuencias tenían una alta rigidez de estiramiento.

5. Estudiamos el efecto de la secuencia de nucleótidos en el doblamiento del dsRNA. Nuestras simulaciones de MD predijeron que secuencias de adeninas y uracilos alternados, que llamamos *AU-tracts*, comprimirían el

surco mayor del dsRNA, dando lugar a una hélice doblada. Estos *AU-tracts* tendrían que ser de al menos tres pares de bases, aunque la extensión óptima son cuatro pares de bases. De cara a confirmar experimentalmente esta predicción, tomamos imágenes de AFM de moléculas de dsRNA que contenían repeticiones periódicas de *AU-tracts*. Observamos que estas moléculas se doblaban más que moléculas de dsRNA de secuencia arbitraria, en consonancia con las predicciones computacionales.

7.4 Perspectivas Futuras

Los resultados que hemos presentado en esta Tesis motivan un estudio más detallado de determinados temas. A continuación qué puntos podrían ser abordados en mayor profundidad en un futuro próximo.

- En el Capítulo 3 mostramos que la hipermetilación de la secuencia poly-CG aumentaba la flexibilidad de estiramiento del dsDNA. Sin embargo, trabajos previos han observado que la metilación de los dinucleótidos CpG confieren al dsDNA una mayor rigidez en el contexto de ciclación y formación de nucleosomas [142]. Existen dos posibles explicaciones para esta discrepancia. La primera es que la metilación afecta de forma diferente la flexibilidad de estiramiento y la de doblamiento. La otra solución es que la secuencia juega un papel en el impacto de la metilación sobre la flexibilidad del dsDNA. En este caso, el contexto en el que se encuentran los dinucleótidos CpG, así como la distribución de las citosinas metiladas podrían determinar si la metilación aumenta o reduce la flexibilidad del dsDNA.

- En el Capítulo 5, la respuesta mecánica de las moléculas con *A-tracts* mostraron una clara desviación del modelo WLC a fuerzas menores que 10 pN. Hasta el momento, no hemos encontrado un modelo teórico que describa de forma satisfactoria estos datos experimentales. En el caso más sencillo en el que no hay fuerza externa, la solución analítica del IBWLC fue altamente no trivial. Así que parece improbable encontrar una fórmula analítica para el IBWLC bajo fuerza. Posiblemente el enfoque

más adecuado para abordar este problema sea usar simulaciones de grano grueso.

- En el Capítulo 6 estudiamos cómo los *AU-tracts* inducen un doblamiento en el dsRNA. Nuestras simulaciones de MD y nuestros experimentos de AFM, así como estructuras cristalográficas previas, coincidieron en que los *AU-tracts* doblan la estructura del dsRNA. Sin embargo, este doblamiento no se observó en la electroforesis en gel. Una posible explicación de este conflicto sería que el doblamiento del *AU-tract* sólo aparezca para concentraciones de cationes superiores a un cierto umbral. Y que estas condiciones no se dan en los geles. Curiosamente, este fue el caso de las secuencias G_3C_3 del dsDNA [219]. La estructura cristalográfica mostraba que esta secuencia doblaba el dsDNA [220], sin embargo este efecto sólo pudo ser detectado en geles al emplear una concentración de sal suficientemente elevada [219]. Podría seguirse esta misma estrategia para estudiar en geles el doblamiento inducido por *AU-tracts* en el dsRNA.

List of publications by author

Publications related to this thesis

- **A. Marin-Gonzalez***, J. G. Vilhena*, R. Perez, F. Moreno-Herrero "Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics", Proc. Natl. Acad. Sci. USA **114**, 7049 (2017).

- **A. Marin-Gonzalez***, J. G. Vilhena*, F. Moreno-Herrero, R. Perez, "DNA Crookedness Regulates DNA Mechanical Properties at Short Length Scales", Phys. Rev. Lett. **122**, 048102 (2019). *(Highlighted in "DNA's crooked path", Nat. Rev. Phys. **1**, 103 (2019).)*

- **A. Marin-Gonzalez**, J. G. Vilhena, F. Moreno-Herrero, R. Perez, "Sequence-dependent mechanical properties of double-stranded RNA", Nanoscale **11**, 21417 (2019).

- **A. Marin-Gonzalez**, C. L. Pastrana, R. Bocanegra, A. Martin-Gonzalez, J. G. Vilhena, R. Perez, B. Ibarra, C. Aicart-Ramos, F. Moreno-Herrero, "Understanding the paradoxical mechanical response of in-phase A-tracts at different force regimes", bioRxiv, doi:10.1101/854968

- **A. Marin-Gonzalez**, C. Aicart-Ramos, M. Marin-Baquero, A. Martin-Gonzalez, M. Suomalainen, A. Kannan, J. G. Vilhena, U. F. Greber, F. Moreno-Herrero, R. Perez, "AU-tracts induce sequence-dependent bending in the RNA duplex", *In preparation.*

*: equally contributing

Appendices

A Details on Constant-Force Molecular Dynamics

Our constant-force molecular dynamics (CFMD) simulations were performed using a distance restrain that is incorporated in the latest versions of the Amber software [76] package. This restrain function is defined by six parameters: r_1, r_2, r_3, r_4, rk_2, and rk_3, where $r_1 < r_2 < r_3 < r_4$ and $rk_2, rk_3 > 0$. The $r_i's$ split the domain of the restrain coordinate in five regions, see Fig. 1.7. The restraining potential, $U(R)$ is given by:

$$U(R) = \begin{cases} A_1 R + B_1 & \text{if } R \leq r_1 \\ rk_2(R - r_2)^2 & \text{if } r_1 < R \leq r_2 \\ 0 & \text{if } r_2 < R \leq r_3 \\ rk_3(R - r_3)^2 & \text{if } r_3 < R \leq r_4 \\ A_2 R + B_2 & \text{if } R > r_4. \end{cases} \tag{A.1}$$

where R is the distance between the atoms on which the restrain acts, which in our case are the virtual atoms described in Section 1.5.1 (purple beads in Fig. 1.6). A_1, A_2, B_1, and B_2 are not free parameters, but are univocally defined by the $r_i's$ and $rk_i's$ and the condition of continuity in $U(R)$ and its derivative. Remind from Section 1.5.1 that in order to have a constant stretching force, our restraining coordinate should only explore the region where $U(R)$ is linear with negative slope, that is, $R \leq r_1$ (blue region of Fig. 1.7). Therefore, the parameters A_2 and B_2 are irrelevant in our case. In addition, because the stretching force is obtained by differentiating $U(R)$, the constant B_1 will not appear in the expression of the force. Hence, we only need to determine

the parameter A_1. This can be done from the condition of continuity of the differential of $U(R)$ at r_1

$$\left(\frac{dU}{dR}\right)_{R=r_1^+} \stackrel{!}{=} \left(\frac{dU}{dR}\right)_{R=r_1^-} \Rightarrow A_1 = 2rk_2(r_1 - r_2) = -2rk_2(r_2 - r_1) \quad \text{(A.2)}$$

Notice that, since $rk_2 > 0$ and $r_2 > r_1$, the equation above implies that $A_1 < 0$, confirming the anticipated statement that the slope of $U(R)$ is negative in the $R \leq r_1$ region. The force is obtained by differentiating the potential:

$$\vec{F} = -\vec{\nabla}U(R) = -\frac{dU(R)}{dR} = -A_1\vec{u}_R = 2rk_2(r_2 - r_1)\vec{u}_R \quad \text{(A.3)}$$

where $\vec{u}_R$ is a vector along the direction of the line that joins the restrained atoms. Equation A.3 illustrates that the force *i)* is constant in modulus, *ii)* always has the direction of the line joining the restrained atoms ($\vec{u}_R$) and *iii)* is positive, meaning that it points outwards the two restrained atoms. Altogether, Eq. A.3 states that we have a constant stretching force.

We implemented our CFMD simulations by choosing an arbitrarily large value of r_1, to ensure that the restraining coordinate remains in the $R < r_1$ during the entire simulation trajectory. We also fixed the value of $r_2 = r_1$+0.5, so that Eq. A.2 becomes $A_1 = -rk_2$. Taking into account that rk_2 is given in kcal mol^{-1} Å^{-2} and the $r_i's$ are in Å, one needs to include a conversion factor of 69.5. All things considered, a representative input restrain of a constant 20 pN force takes the form:

```
&rst iat= -1, -1,
    r1 = 100.5 , r2 = 101, r3 = 102 , r4 = 103 ,
    rk2 = 0.2877697 , rk3 =0.0001 ,
    igr1 = 42, 959, 0, igr2 = 551, 456, 0,
&end
```

B Constitutive Equations of the Elastic Rod Model

In the regime where bending fluctuations are negligible, the elastic energy of a NA can be written as [37, 42, 43, 100] (Section 1.3.1)

$$E(x,\theta,F) = \frac{1}{2}\frac{C}{L}\theta^2 + g\theta\frac{x}{L} + \frac{1}{2}\frac{S}{L}x^2 - xF \tag{B.1}$$

where L is the equilibrium extension at zero force, x is the elongation or deviation from L and θ is the change in helical twist from its unperturbed equilibrium value. The three parameters S, C, g and are the stretch modulus, the twist rigidity, and the twist-stretch coupling, respectively.

If F and θ are fixed (F_0, θ_0), the value of the extension (x^*) that minimizes the energy is given by

$$\left(\frac{\partial E}{\partial x}\right)_{F=F_0,\theta=\theta_0} = 0 \;\Rightarrow\; g\frac{\theta_0}{L} + \frac{S}{L}x^* - F_0 = 0 \;\Rightarrow\; x^* = \frac{L}{S}\left(F_0 - \frac{g\theta_0}{L}\right) \tag{B.2}$$

By the same token, at fixed F and x

$$\left(\frac{\partial E}{\partial \theta}\right)_{F=F_0,x=x_0} = 0 \;\Rightarrow\; gx_0 + C\theta^* = 0 \;\Rightarrow\; \theta^* = -\frac{g}{C}\,x_0 \tag{B.3}$$

Moreover, at a given value of the force there is a global minimum of the energy in (x,θ). We denote this minimum $(<x>,<\theta>)$. We can take $\theta_0 =<\theta>$ in Eq. B.2 and $x_0 =<x>$ in Eq. B.3, obtaining

$$<\theta> = -\frac{g}{C}<x> \tag{B.4}$$

$$\frac{<x>}{L} = \frac{F}{S} - g\,\frac{<\theta>}{SL} \tag{B.5}$$

By substituting the value of $< \theta >$ from Eq. B.4 in Eq. B.5, one gets [100]

$$\frac{< x >}{L} = \frac{F}{S - g^2/C} \equiv \frac{F}{\tilde{S}} \tag{B.6}$$

As pointed out in [100], $S > \tilde{S} = S - g^2/C$, i.e. allowing the molecule to rotate softens the stretching elasticity.

Substituting the $< x >$ from Eq. B.6 in Eq. B.4 yields for the helical twist

$$\frac{< \theta >}{L} = -\frac{g}{C\tilde{S}} F \tag{B.7}$$

Finally, a third relation can be obtained if the twisting angle imposed in Eq. B.2 is changed while keeping the force constant. The equilibrium extension then varies as [37, 42, 43]

$$\left(\frac{\partial x^*}{\partial \theta_0}\right)_F = -\frac{g}{S} \tag{B.8}$$

Hence, in order to compute the parameters of the elastic rod model one needs to measure three quantities: the dependence of the elongation (Eq. B.6) and the twisting angle (Eq. B.7) with the force and how thermal fluctuations of the twisting angle are coupled to fluctuations of the elongation (Eq. B.8). Notice that the equations are coupled, so the determination of the elastic parameters requires solving the system of equations.

We denote $A_1 \equiv 1/\tilde{S}$, $A_2 \equiv -g/\tilde{S}C$ and $A_3 \equiv -g/S$, obtaining

$$S = \frac{1}{A_1 - A_2 A_3} \tag{B.9}$$

$$C = \frac{A_1 A_3 / A_2}{A_1 - A_2 A_3} \tag{B.10}$$

$$g = -\frac{A_3}{A_1 - A_2 A_3} \tag{B.11}$$

These equations allow us to compute all elastic parameters from the slopes of Fig. 2.1a, b (A_1, A_2) and from the helical rise - helical twist slopes at F =1 pN obtained from Fig. 2.2 and shown in Fig. 2.1c (A_3).

For the sake of simplicity, $< x >$, x^* and x_0 are all denoted as x in the main text; the same holds for θ.

C Springiness and Crookedness Models

C.1 Springiness Model

Given a dsDNA/dsRNA molecule, one can define a chain of segments that joins the centers of all consecutive base pairs (see Section 2.3). If bending is neglected, the extension (h) of the molecule is equal to

$$h = h(l, \cos\beta) = l\cos\beta, \tag{C.1}$$

where l is the length is the length of the chain and β is the springiness of the system, or the angle defined by the chain with the helical axis. Introducing an external force will induce a variation in these quantities

$$h(l(F), \cos\beta(F)) = l(F)\ \cos\beta(F) \tag{C.2}$$

If the change in h induced by the force is small enough, we can write $h(F)$ as a Taylor expansion around $h(F=1) \equiv L$ and stay at the first order approximation. Denoting $l(F=1) \equiv l_0$, $\cos\beta(F=1) \equiv \cos\beta_0$, we get

$$\begin{aligned} h(F) - h(1) \equiv x &= \left(\frac{\partial h(l,\cos\beta)}{\partial l}\right)_{\substack{l=l_0 \\ \cos\beta=\cos\beta_0}} \times \Delta l \\ &+ \left(\frac{\partial h(l,\cos\beta)}{\partial(\cos\beta)}\right)_{\substack{l=l_0 \\ \cos\beta=\cos\beta_0}} \times \Delta(\cos\beta) \qquad \text{(C.3)} \\ &= \cos\beta_0\ \Delta l + l_0\ \Delta(\cos\beta) \qquad \text{(C.4)} \end{aligned}$$

In the equation above, the first term accounts for the change in extension coming from separating consecutive base pairs from each other. The second term measures the contribution from deforming the centerline polymer, changing its

end-to-end distance when its contour length is kept fixed. Consequently, we define these quantities as

$$x_{\Delta l} \equiv \cos\beta_0 \, \Delta l; \qquad\qquad x_{\Delta\beta} \equiv l_0 \, \Delta(\cos\beta) \tag{C.5}$$

and then

$$x = x_{\Delta l} + x_{\Delta\beta} \tag{C.6}$$

As shown above (Eq. B.6 and Fig. 2.1), in the elastic rod model x varies linearly with the force as $x = (FL)/\tilde{S}$. Figure 2.3 shows that $x_{\Delta l}$ and $x_{\Delta\beta}$ also depend linearly on the force. Therefore, we can define the elastic constants k_l and k_β associated to $x_{\Delta l}$ and $x_{\Delta\beta}$ and express these quantities as

$$\frac{x_{\Delta l}}{L} = \frac{F}{k_l} \qquad\qquad \frac{x_{\Delta\beta}}{L} = \frac{F}{k_\beta} \tag{C.7}$$

Then, from these two relations and Eq. C.6 and B.6, we get

$$\frac{x}{FL} = \frac{1}{\tilde{S}} = \frac{1}{k_l} + \frac{1}{k_\beta} \tag{C.8}$$

or simply

$$\tilde{S} = \frac{k_l k_\beta}{k_l + k_\beta} \tag{C.9}$$

This equation illustrates that we are just modelling the elastic response of NA as two springs in series of elastic constants k_l and k_β.

C.2 Crookedness Model

We will follow a similar reasoning for the crookedness model, but in this case we will express the contour length of the chain as the sum of the distances between consecutive base pairs, l_i. The extension of an N-bp molecule is then equal to

$$h(F) = \cos\beta(F) \times \sum_{i=1}^{N-1} l^{(i)}(F) \tag{C.10}$$

where the summatory extends over the $N-1$ base pair steps. Now we can span $h(F)$ as a Taylor series of an N-dimensional function around the point $(l^{(i)} = l^{(i)}(0), \cos\beta = \cos\beta(0))$. In this case, we get

$$h(F) - h(1) \equiv x = \sum_{i=1}^{N-1} \cos\beta_0 \, \Delta l^{(i)} + \left(\sum_{i=1}^{N-1} l_0^{(i)} \right) \Delta(\cos\beta). \tag{C.11}$$

Then, similar to the springiness model (Eq. C.5), we can define

$$x_{\Delta l^{(i)}} \equiv \cos\beta_0 \; \Delta l^{(i)}; \tag{C.12}$$

$$x_{\Delta\beta} \equiv \left(\sum_{i=1}^{N-1} l_0^{(i)} \right) \; \Delta(\cos\beta). \tag{C.13}$$

The elastic constants associated to these deformations are

$$k_{l,i} \equiv \frac{LF}{x_{\Delta l^{(i)}}} \tag{C.14}$$

$$k_{\beta} \equiv \frac{LF}{x_{\Delta\beta}}. \tag{C.15}$$

Finally, using these definitions, Eq. C.11, and the definition of the effective stretch modulus (Eq. B.6) one arrives at

$$S^{-1} = \sum_{i=1}^{N-1} k_{l,i}^{-1} + k_{\beta}^{-1}. \tag{C.16}$$

Therefore, similar to the springiness case, we are approximating the stretching response of the molecule as a set of springs in series, but in this case there are N springs: $N-1$ corresponding to separating consecutive base pairs $(k_{l,i})$ and an N^{th} spring that accounts for the crookedness deformation (k_{β}). This idea is schematically illustrated in Fig. 3.3.

D Intrinsically Bent Worm-Like Chain

D.1 Worm-Like Chain (WLC) Model

It is helpful to start with the simpler case of the WLC model. Here we will follow the derivation of the model from refs [31, 182]. According to the 2D WLC model, the bending energy of the polymer is given by

$$E = \frac{Pk_BT}{2L}\varphi^2 \tag{D.1}$$

where P is the persistence length; φ is the bending angle between the tangents of two points separated by a distance L; T is the temperature; and k_B the Boltzmann constant. From this expression, the probability of finding a given bending angle is simply

$$\mathcal{P}(\varphi) = \sqrt{\frac{P}{2\pi L}}e^{-\frac{P\varphi^2}{2L}} \tag{D.2}$$

From here one can easily obtain the mean cosine of the bending angle, which is given by the following integral

$$< \cos\varphi >_{WLC} = \int_{-\infty}^{\infty} d\varphi \cos\varphi \mathcal{P}(\varphi) = e^{-L/2P} \tag{D.3}$$

and then the mean squared end-to-end distance

$$\begin{aligned} < R^2 > &= \int_0^L ds \int_0^L ds' < \cos(\varphi(s,s')) > = 2\int_0^L ds \int_0^s ds' e^{\frac{-(s-s')}{2P}} \\ &= 4PL\left[1 - \frac{2P}{L}(1 - e^{-L/2P})\right] \end{aligned} \tag{D.4}$$

D.2 Intrinsically-Bent WLC (IBWLC) Model

For an intrinsically bent molecule the minimum energy does not correspond to a zero bending angle. We will assume that the intrinsic bending describes a circle of radius $R_0 = 1/a$, such that the energy reads

$$E = \frac{Pk_BT}{2L}(\varphi - aL)^2 \tag{D.5}$$

Then, the angle probability is

$$\mathcal{P}(\varphi) = \sqrt{\frac{P}{2\pi L}} e^{\frac{-P(\varphi - aL)}{2L}} \tag{D.6}$$

And the mean cosine can be obtained as

$$\begin{aligned} < \cos\varphi > &= \sqrt{\frac{P}{2\pi L}} \int_{-\infty}^{\infty} d\varphi \cos\varphi e^{\frac{-P(\varphi - aL)^2}{2L}} \\ &= \sqrt{\frac{P}{2\pi L}} \int_{-\infty}^{\infty} d\varphi \cos(\varphi + aL) e^{\frac{-P\varphi^2}{2L}} \\ &= \cos(aL) e^{-\frac{L}{2P}} \end{aligned} \tag{D.7}$$

This expression is shown in Chapter 5 and was used to fit the A-tracts AFM data. The formula for the mean squared end-to-end distance can be derived in the same way as done for the WLC. In this case, one would need to solve the following integral

$$< R^2 >= 2 \int_0^L ds \int_0^s ds' \cos(a(s - s')) e^{\frac{-(s-s')}{2P}} \tag{D.8}$$

This integral is more cumbersome than for the WLC case, but can also be solved analytically yielding the expression shown in the main text

$$\begin{aligned} < R^2 > = &\frac{2}{a^2 + b^2} \times \\ &\times \left\{ bL + \frac{1}{a^2 + b^2} \left[(a^2 - b^2)(1 - \cos(aL) e^{-bL}) - 2ab \sin(aL) e^{-bL} \right] \right\} \end{aligned} \tag{D.9}$$

with $b \equiv 1/(2P_0)$. Notice that by making $a = 0$ one recovers the expression of the WLC (Eq. D.4) as required.

D.3 A More Elegant Derivation of the IBWLC Model

A more elegant solution to the problem can be found by working in the complex plane. Realize that

$$< \cos\varphi >= \mathcal{R}e\left\{e^{-\frac{L}{2P^*}}\right\} \tag{D.10}$$

where $\mathcal{R}e$ denotes the real part and P^* is a complex number given by

$$P^* = \frac{P}{1-2iPa} \tag{D.11}$$

The mean squared end-to-end distance is then obtained in a straightforward way, building upon the results obtained for the WLC

$$\begin{aligned} < R^2 > &= \int_0^L ds \int_0^L ds' \mathcal{R}e\left\{e^{-\frac{|s-s'|}{2P^*}}\right\} = \mathcal{R}e\left\{\int_0^L ds \int_0^L ds' e^{-\frac{|s-s'|}{2P^*}}\right\} \\ &= \mathcal{R}e\left\{4P^*L\left[1-\frac{2P^*}{L}\left(1-e^{-\frac{L}{2P^*}}\right)\right]\right\} \end{aligned} \tag{D.12}$$

This expression reveals that the mean squared end-to-end distance in the IBWLC is simply the real part of the one from the WLC when a complex persistence length is used. We will express this as

$$< R^2_{IBWLC} >= \mathcal{R}e < R^2_{WLC;P\to P^*} > \tag{D.13}$$

Expanding Eq. D.12 one arrives at Eq. D.9.

Bibliography

[1] O. T. Avery, C. M. Macleod, and M. McCarty, J. Exp. Med. **79**, 137 (1944).

[2] A. D. Hershey and M. Chase, J. Gen. Physiol. **36**, 39 (1952).

[3] R. E. Franklin and R. G. Gosling, Nature **171**, 740 (1953).

[4] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, Nature **171**, 738 (1953).

[5] E. Chargaff, Experientia **6**, 201 (1950).

[6] J. D. Watson and F. H. C. Crick, Nature **171**, 737 (1953).

[7] F. H. C. Crick, *What Mad Pursuit. A Personal View of Scientific Discovery* (BasicBooks, 1988).

[8] P. Lengyel, J. F. Speyer, and S. Ochoa, Proc. Natl. Acad. Sci. U.S.A. **47**, 1936 (1961).

[9] M. R. Bernfield and M. W. Nirenberg, Science **147**, 479 (1965).

[10] M. Egli, A. Flavell, A. M. Pyle, W. D. Wilson, S. I. Haq, B. Luisi, J. Fisher, C. Laughton, S. Allen, and J. Engels, *Nucleic Acids in Chemistry and Biology*, edited by G. M. Blackburn, M. J. Gait, D. Loakes, and D. M. Williams (The Royal Society of Chemistry, 2006).

[11] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, Nature **391**, 806 (1998).

[12] M. Yoneyama, M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira, and T. Fujita, Nat. Immunol. **5**, 730 (2004).

[13] Y.-H. Chiu, J. B. MacMillan, and Z. J. Chen, Cell **138**, 576 (2009).

[14] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan, Nature **407**, 327 (2000).

[15] P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, and T. A. Steitz, Proc. Natl. Acad. Sci. U.S.A. **98**, 4899 (2001).

[16] R. R. Sinden, *DNA Structure and Function* (Academic Press, 1994).

[17] V. Bloomfield, P. Killman, D. Crothers, I. Tinoco, J. Hearst, D. Wemmer, and D. Turner, *Nucleic Acids: Structure, Properties, and Functions* (University Science Books, 2000).

[18] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).

[19] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, Nature **287**, 755 (1980).

[20] M. A. E. Hassan and C. R. Calladine, Philos. Trans. R. Soc. A **355**, 43 (1997).

[21] C. A. Hunter, J. Mol. Biol. **230**, 1025 (1993).

[22] M. J. Packer and C. A. Hunter, J. Mol. Biol. **280**, 407 (1998).

[23] J. Sponer, G. Bussi, M. Krepl, P. Banás, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurecka, N. G. Walter, and M. Otyepka, Chem. Rev. **118**, 4177 (2018).

[24] M. McCall, T. Brown, and O. Kennard, J. Mol. Biol. **183**, 385 (1985).

[25] H. C. M. Nelson, J. T. Finch, B. F. Luisi, and A. Klug, Nature **330**, 221 (1987).

[26] H.-S. Koo, H.-M. Wu, and D. M. Crothers, Nature **320**, 501 (1986).

[27] X. Lu and W. K. Olson, Nucleic Acids Res. **31**, 5108 (2003).

[28] F. Kriegel, N. Ermann, and J. Lipfert, J. Struct. Biol. **197**, 26 (2017).

[29] S. K. Nomidis, F. Kriegel, W. Vanderlinden, J. Lipfert, and E. Carlon, Phys. Rev. Lett. **118**, 217801 (2017).

[30] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, Nat. Nanotechnol. **1**, 137 (2006).

[31] C. Rivetti, M. Guthold, and C. Bustamante, J. Mol. Biol. **264**, 919 (1996).

[32] E. Herrero-Galán, M. E. Fuentes-Perez, C. Carrasco, J. M. Valpuesta, J. L. Carrascosa, F. Moreno-Herrero, and J. R. Arias-Gonzalez, J. Am. Chem. Soc. **135**, 122 (2013).

[33] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante, Annu. Rev. Biochem. **77**, 205 (2008).

[34] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).

[35] C. Bustamante, J. Marko, E. Siggia, and S. Smith, Science **265**, 1599 (1994).

[36] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).

[37] J. Lipfert, G. M. Skinner, J. M. Keegstra, T. Hensgens, T. Jager, D. Dulin, M. Köber, Z. Yu, S. P. Donkers, F.-c. Chou, R. Das, and N. H. Dekker, Proc. Natl. Acad. Sci. U.S.A. **111**, 15408 (2014).

[38] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature **424**, 338 (2003).

[39] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, Nat. Methods **7**, 977 (2010).

[40] F. Kriegel, N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, and J. Lipfert, Nucleic Acids Res. **45**, 5920 (2017).

[41] J. Lipfert, M. Wiggin, J. W. J. Kerssemakers, F. Pedaci, and N. H. Dekker, Nature Commun. **2**, 439 (2011).

[42] J. Gore, Z. Bryant, M. Nöllmann, M. U. Le, N. R. Cozzarelli, and C. Bustamante, Nature **442**, 836 (2006).

[43] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **96**, 1 (2006).

[44] P. Lebel, A. Basu, F. C. Oberstrass, E. M. Tretter, and Z. Bryant, Nat. Methods **11**, 456 (2014).

[45] P. Gross, N. Laurens, L. B. Oddershede, U. Bockelmann, E. J. G. Peterman, and G. J. L. Wuite, Nat. Phys. **7**, 731 (2011).

[46] O. D. Broekmans, G. A. King, G. J. Stephens, and G. J. L. Wuite, Phys. Rev. Lett. **116**, 258102 (2016).

[47] W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin, Proc. Nat. Acad. Sci. U.S.A. **95**, 11163 (1998).

[48] F. C. Chou, J. Lipfert, and R. Das, PLOS Comput. Biol. **10**, 37 (2014).

[49] D. L. Beveridge *et al.*, Biophys. J. **87**, 3799 (2004).

[50] S. B. Dixit *et al.*, Biophys. J. **89**, 3721 (2005).

[51] R. Lavery *et al.*, Nucleic Acids Res. **38**, 299 (2009).

[52] M. Pasi *et al.*, Nucleic Acids Res. **42**, 12272 (2014).

[53] A. K. Mazur, J. Phys. Chem. B **113**, 2077 (2009).

[54] A. Garai, S. Saurabh, Y. Lansac, and P. K. Maiti, J. Phys. Chem. B **119**, 11146 (2015).

[55] Y.-Y. Wu, L. Bao, X. Zhang, and Z.-J. Tan, J. Chem. Phys. **142**, 125103 (2015).

[56] F. Lankas, J. Sponer, P. Hobza, and J. Langowski, J. Mol. Biol. **299**, 695 (2000).

[57] A. K. Mazur, Phys. Rev. E **84**, 021903 (2011).

[58] I. Faustino, A. Pérez, and M. Orozco, Biophys. J. **99**, 1876 (2010).

[59] A. K. Mazur, Biophys. J. **91**, 4507 (2006).

[60] A. Noy and R. Golestanian, Phys. Rev. Lett. **109**, 228101 (2012).

[61] A. Noy, A. Pérez, F. Lankas, F. J. Luque, and M. Orozco, J. Mol. Biol. **343**, 627 (2004).

[62] T. Drsata, N. Špačková, P. Jurečka, M. Zgarbová, J. Sponer, and F. Lankas, Nucleic Acids Res. **42**, 7383 (2014).

[63] F. Lankaš, T. E. Cheatham, N. Špačáková, P. Hobza, J. Langowski, and J. Šponer, Biophys. J. **82**, 2592 (2002).

[64] A. K. Mazur, J. Chem. Theory Comput. **5**, 2149 (2009).

[65] L. Bao, X. Zhang, Y.-Z. Shi, Y.-Y. Wu, and Z.-J. Tan, Biophys. J. **112**, 1094 (2017).

[66] F. Lankaš, J. Sponer, J. Langowski, and T. E. Cheatham, Biophys. J. **85**, 2872 (2003).

[67] A. K. Mazur, Phys. Rev. E **86**, 011914 (2012).

[68] A. Reymer, K. Zakrzewska, and R. Lavery, Nucleic Acids Res. **46**, 1684 (2017).

[69] A. K. Mazur, J. Phys. Chem. B **117**, 1857 (2013).

[70] F. Lankas, N. Speckova, M. Moakher, P. Enkhbayar, and J. Sponer, Nucleic Acids Res. **38**, 3414 (2010).

[71] A. V. Drozdetski, I. S. Tolokh, L. Pollack, N. Baker, and A. V. Onufriev, Phys. Rev. Lett. **117**, 028101 (2016).

[72] K. Liebl, T. Drsata, F. Lankas, J. Lipfert, and M. Zacharias, Nucleic Acids Res. **43**, 10143 (2015).

[73] T. Lionnet and F. Lankaš, Biophys. J. **92**, L30 (2007).

[74] I. Besseova, M. Otyepka, K. Réblová, and J. Sponer, Phys. Chem. Chem. Phys. **11**, 10701 (2009).

[75] I. Besseová, P. Banás, P. Kührová, P. Kosinová, M. Otyepka, and J. Sponer, J. Phys. Chem. B **116**, 9899 (2012).

[76] D. Case *et al.*, "Amber 14," (2014).

[77] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **9**, 3878 (2013).

[78] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **8**, 1542 (2012).

[79] S. L. Grand, A. W. Götz, and R. C. Walker, Comput. Phys. Commun. **184**, 374 (2013).

[80] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. C. III, C. A. Laughton, and M. Orozco, Biophys. J. **92**, 3817 (2007).

[81] M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T. E. Cheatham, and P. Jurecka, J. Chem. Theory Comput. **7**, 2886 (2011), pMID: 21921995.

[82] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).

[83] I. S. Joung and T. E. Cheatham, J. Phys. Chem. B **113**, 13279 (2009).

[84] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[85] K. M. Kosikov, A. A. Gorin, V. B. Zhurkin, and W. K. Olson, J. Mol. Biol. **289**, 1301 (1999).

[86] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Res **37**, 5917 (2009).

[87] R. Garcia and R. Perez, Surf. Sci. Rep. **47**, 197 (2002).

[88] I. Horcas, R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, and A. M. Baro, Rev. Sci. Instrum. **78**, 013705 (2007).

[89] F. Moreno-Herrero, R. Seidel, S. M. Johnson, A. Fire, and N. H. Dekker, Nucleic Acids Res. **34**, 3057 (2006).

[90] C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003).

[91] K. M. Kosikov, A. A. Gorin, V. B. Zhurkin, and W. K. Olson, J. Mol. Biol. **289**, 1301 (1999).

[92] N. R. Cozzarelli, G. J. Cost, M. Nollmann, T. Viard, and J. E. Stray, Nat. Rev. Mol. Cell Biol. **7**, 580 (2006).

[93] R. Schleif, Annu. Rev. Biochem. **61**, 199 (1992).

[94] A. B. Conway, T. W. Lynch, Y. Zhang, G. S. Fortin, C. W. Fung, L. S. Symington, and P. A. Rice, Nat Struct Mol Biol **11**, 791 (2004).

[95] J. T. P. Yeeles, K. van Aelst, M. S. Dillingham, and F. Moreno-Herrero, Mol. Cell **42**, 806 (2011).

[96] D. Dulin, I. D. Vilfan, B. A. Berghuis, M. M. Poranen, M. Depken, and N. H. Dekker, Nucleic Acids Res. **43**, 10421 (2015).

[97] I. D. Vlaminck and C. Dekker, Annu. Rev. Biophys. **41**, 453 (2012).

[98] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante, Annu. Rev. Biochem. **77**, 205 (2008).

[99] J. A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker, and N. H. Dekker, Biophys. J. **88**, 2737 (2005).

[100] J. F. Marko, EPL **38**, 183 (2007).

[101] P. Nelson, *Biological Physics* (Freeman, W. H. and Company, New York, 2007).

[102] A. Pérez, F. J. Luque, and M. Orozco, J. Am. Chem. Soc. **129**, 14739 (2007).

[103] A. Pérez, F. J. Luque, and M. Orozco, Acc. Chem. Res. **45**, 196 (2012).

[104] P. D. Dans, A. Perez, I. Faustino, R. Lavery, and M. Orozco, Nucleic Acids Res. **40**, 10668 (2012).

[105] R. Galindo-Murillo, D. R. Roe, and T. E. Cheatham, Nature Commun. **5**, 5152 (2014).

[106] S. Kannan, K. Kohlhoff, and M. Zacharias, Biophys. J. **91**, 2956 (2006).

[107] B. Luan and A. Aksimentiev, Phys. Rev. Lett. **101**, 1 (2008).

[108] J. Rezac, P. Hobza, and S. A. Harris, Biophys. J. **98**, 101 (2010).

[109] L. Bongini, V. Lombardi, and P. Bianco, J. Royal Soc. Interface **11**, 20140399 (2014).

[110] P. M. D. Severin, X. Zou, H. E. Gaub, and K. Schulten, Nucleic Acids Res. **39**, 8740 (2011).

[111] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, Biophys. J. **82**, 3160 (2002).

[112] M. Y. Sheinin and M. D. Wang, Phys. Chem. Chem. Phys. **11**, 4800 (2009).

[113] J. D. Moroz and P. Nelson, Macromolecules **31**, 6333 (1998).

[114] J. M. Vargason, K. Henderson, and P. S. Ho, Proc. Natl. Acad. Sci. U.S.A. **98**, 7265 (2001).

[115] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, Annu. Rev. Biochem. **79**, 233 (2010).

[116] X.-J. Lu, Z. Shakked, and W. K. Olson, J. Mol. Biol. **300**, 819 (2000).

[117] S. Ahmad, H. Kono, M. J. Arauzo-Bravo, and A. Sarai, Nucleic Acids Res. **34**, W124 (2006).

[118] M. H. Werner, A. M. Gronenborn, and G. M. Clore, Science **271**, 778 (1996).

[119] T. E. Haran and U. Mohanty, Q. Rev. Biophys. **42**, 41–81 (2009).

[120] M. McCall, T. Brown, W. N. Hunter, and O. Kennard, Nature **322**, 661 (1986).

[121] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig, Nature **461**, 1248 (2009).

[122] E. Segal and J. Widom, Curr. Opin. Struct. Biol. **19**, 65 (2009).

[123] A. M. Deaton and A. Bird, Genes Dev. **25**, 1010 (2011).

[124] H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins, Biopolymers **85**, 115 (2005).

[125] A. Marín-González, J. Vilhena, R. Pérez, and F. Moreno-Herrero, Proc. Natl. Acad. Sci. U.S.A **114**, 7049 (2017).

[126] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, Nature **287**, 755 EP (1980).

[127] J. L. Kim, D. B. Nikolov, and S. K. Burley, Nature **365**, 520 (1993).

[128] D. Bramhill and A. Kornberg, Cell **52**, 743 (1988).

[129] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond, Nature **389**, 251 (1997).

[130] K. E. Flick, M. S. Jurica, R. J. Monnat Jr, and B. L. Stoddard, Nature **394**, 96 (1998).

[131] S. Fujii, H. Kono, S. Takenaka, N. Go, and A. Sarai, Nucleic Acids Res. **35**, 6063 (2007).

[132] W. K. Olson, A. V. Colasanti, Y. Li, W. Ge, G. Zheng, and V. B. Zhurkin, *Computational studies of RNA and DNA*, edited by J. Sponer and F. Lankas (Springer, 2006) Chap. 9, pp. 235–257.

[133] F. Kilchherr, C. Wachauf, B. Pelz, M. Rief, M. Zacharias, and H. Dietz, Science **353** (2016), 10.1126/science.aaf5508.

[134] C. I. Pongor, P. Bianco, G. Ferenczy, R. Kellermayer, and M. Kellermayer, Biophys. J. **112**, 512 (2017).

[135] F. Lankas, Biopolymers **73**, 327 (2004).

[136] R. Vlijm, J. v.d. Torre, and C. Dekker, PLOS ONE **10**, 1 (2015).

[137] S. H. Kim, M. Ganji, E. Kim, J. van der Torre, E. Abbondanzieri, and C. Dekker, eLife **7**, e36557 (2018).

[138] V. R. Ramirez-Carrozzi, D. Braas, D. M. Bhatt, C. S. Cheng, C. Hong, K. R. Doty, J. C. Black, A. Hoffmann, M. Carey, and S. T. Smale, Cell **138**, 114 (2009).

[139] S. B. Baylin, Nat. Clin. Prac. Oncol. **2**, S4 (2005).

[140] A. Perez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. Ruiz, D. Torrents, R. Eritja, M. Soler-López, and M. Orozco, Biophys. J. **102**, 2140 (2012).

[141] G. Portella, F. Battistini, and M. Orozco, PLOS Comput. Biol. **9**, 1 (2013).

[142] T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, and T. Ha, Nat. Commun. **7**, 10813 (2016).

[143] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, Nature **486**, 496 (2012).

[144] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal, Nature **458**, 362 (2009).

[145] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom, Nature **442**, 772 (2006).

[146] H. Qu, C.-Y. Tseng, Y. Wang, A. J. Levine, and G. Zocchi, EPL **90**, 18003 (2010).

[147] R. Vafabakhsh and T. Ha, Science **337**, 1097 (2012).

[148] F. Lankaš, R. Lavery, and J. H. Maddocks, Structure **14**, 1527 (2006).

[149] J. Curuksu, M. Zacharias, R. Lavery, and K. Zakrzewska, Nucleic Acids Res **37**, 3766 (2009).

[150] D. P. Bartel, Cell **116**, 281 (2004).

[151] D. P. Bartel, Cell **136**, 215 (2009).

[152] M. C. Siomi, K. Sato, D. Pezic, and A. A. Aravin, Nat. Rev. Mol. Cell Biol. **12**, 246 (2011).

[153] L. R. Ganser, M. L. Kelly, D. Herschlag, and H. M. Al-Hashimi, Nat. Rev. Mol. Cell Biol. **20**, 474 (2019).

[154] J. E. Wilusz, H. Sunwoo, and D. L. Spector, Genes Dev. **23**, 1494 (2009).

[155] I. Tinoco and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).

[156] P. Brion and E. Westhof, Annu. Rev. Biophys. Biomol. Struct. **26**, 113 (1997).

[157] J. Stombaugh, C. L. Zirbel, E. Westhof, and N. B. Leontis, Nucleic Acids Res. **37**, 2294 (2009).

[158] J. D. Yesselman, S. K. Denny, N. Bisaria, D. Herschlag, W. J. Greenleaf, and R. Das, Proc. Natl. Acad. Sci. U.S.A. **116**, 16847 (2019).

[159] G. Lee, M. A. Bratkowski, F. Ding, A. Ke, and T. Ha, Science **336**, 1726 (2012).

[160] G. Masliah, P. Barraud, and F. H. T. Allain, Cell. Mol. Life Sci. **70**, 1875 (2013).

[161] X. Zheng and P. C. Bevilacqua, Proc. Natl. Acad. Sci. U.S.A. **97**, 14162 (2000).

[162] A. Marin-Gonzalez, J. G. Vilhena, F. Moreno-Herrero, and R. Perez, Phys. Rev. Lett. **122**, 048102 (2019).

[163] A. Pérez, A. Noy, F. Lankas, F. J. Luque, and M. Orozco, Nucleic Acids Res. **32**, 6144 (2004).

[164] K. J. Dechering, K. Cuelenaere, R. N. Konings, and J. A. Leunissen, Nucleic Acids Res. **26**, 4056 (1998).

[165] M. J. Behe, Nucleic Acids Res. **23**, 689 (1995).

[166] M. Y. Tolstorukov, K. M. Virnik, S. Adhya, and V. B. Zhurkin, Nucleic Acids Res. **33**, 3907 (2005).

[167] A. Fire, R. Alcazar, and F. Tan, Genetics **173**, 1259 (2006).

[168] Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal, PLOS Comput. Biol. **4**, 1 (2008).

[169] T. Raveh-Sadka, M. Levo, U. Shabi, B. Shany, L. Keren, M. Lotan-Pompan, D. Zeevi, E. Sharon, A. Weinberger, and E. Segal, Nat. Genet. **44**, 743 (2012).

[170] P. J. Hagerman, Annu. Rev. Biochem. **59**, 755 (1990).

[171] A. Tubbs *et al.*, Cell **174**, 1127 (2018).

[172] A. Ablasser, F. Bauernfeind, G. Hartmann, E. Latz, K. A. Fitzgerald, and V. Hornung, Nat. Immunol. **10**, 1065 (2009).

[173] C. Frøkjær-Jensen, N. Jain, L. Hansen, M. W. Davis, Y. Li, D. Zhao, K. Rebora, J. R. Millet, X. Liu, S. K. Kim, D. Dupuy, E. M. Jorgensen, and A. Z. Fire, Cell **166**, 343 (2016).

[174] E. Stellwagen, Q. Dong, and N. C. Stellwagen, Biophys. J. **108**, 2291 (2015).

[175] L. E. Ulanovsky and E. N. Trifonov, Nature **326**, 720 (1987).

[176] A. Bolshoy, P. McNamara, R. E. Harrington, and E. N. Trifonov, Proc. Natl. Acad. Sci. U.S.A. **88**, 2312 (1991).

[177] A. D. DiGabriele, M. R. Sanderson, and T. A. Steitz, Proc. Natl. Acad. Sci. U.S.A. **86**, 1816 (1989).

[178] A. D. DiGabriele and T. A. Steitz, J. Mol. Biol. **231**, 1024 (1993).

[179] A. A. Podtelezhnikov, C. Mao, N. C. Seeman, and A. Vologodskii, Biophys. J. **79**, 2692 (2000).

[180] Y. Zhang, Z. Xi, R. S. Hegde, Z. Shakked, and D. M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **101**, 8337 (2004).

[181] M. Hogan, J. LeGrange, and B. Austin, Nature **304**, 752 (1983).

[182] C. Rivetti, C. Walker, and C. Bustamante, J. Mol. Biol. **280**, 41 (1998).

[183] S. Johnson, Y.-J. Chen, and R. Phillips, PLOS ONE **8**, 1 (2013).

[184] N. Rezaei, A. Lyons, and N. R. Forde, Biophys. J. **115**, 1457 (2018).

[185] H. S. Koo and D. M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **85**, 1763 (1988).

[186] C. Bouchiat, M. Wang, J.-F. Allemand, T. Strick, S. Block, and V. Croquette, Biophys. J. **76**, 409 (1999).

[187] J.-L. Barrat and J.-F. Joanny, EPL **24**, 333 (1993).

[188] X. Shui, L. McFail-Isom, G. G. Hu, and L. D. Williams, Biochemistry **37**, 8341 (1998).

[189] X. Shui, C. C. Sines, L. McFail-Isom, D. VanDerveer, and L. D. Williams, Biochemistry **37**, 16877 (1998).

[190] K. Tóth, V. Sauermann, and J. Langowski, Biochemistry **37**, 8173 (1998).

[191] S. D. Levene, H. M. Wu, and D. M. Crothers, Biochemistry **25**, 3988 (1986).

[192] E. Stellwagen, J. P. Peters, L. J. Maher, and N. C. Stellwagen, Biochemistry **52**, 4138 (2013).

[193] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).

[194] T. Odijk, Macromolecules **28**, 7016 (1995).

[195] P. A. Wiggins, R. Phillips, and P. C. Nelson, Phys. Rev. E **71**, 021909 (2005).

[196] S. C. Satchwell, H. R. Drew, and A. A. Travers, J. Mol. Biol. **191**, 659 (1986).

[197] T. J. Richmond and C. A. Davey, Nature **423**, 145 (2003).

[198] M. Y. Tolstorukov, A. V. Colasanti, D. M. McCandlish, W. K. Olson, and V. B. Zhurkin, J. Mol. Biol. **371**, 725 (2007).

[199] Y. Bao, C. L. White, and K. Luger, J. Mol. Biol. **361**, 617 (2006).

[200] B. Tian, P. C. Bevilacqua, A. Diegelman-Parente, and M. B. Mathews, Nat. Rev. Mol. Cell Biol. **5**, 1013 (2004).

[201] A. Dock-Bregeon, B. Chevrier, A. Podjarny, J. Johnson, J. de Bear, G. Gough, P. Gilham, and D. Moras, J. Mol. Biol. **209**, 459 (1989).

[202] A. C. Dock-Bregeon, B. Chevrier, A. Podjarny, D. Moras, J. S. deBear, G. R. Gough, P. T. Gilham, and J. E. Johnson, Nature **335**, 375 (1988).

[203] F. U. Gast and P. J. Hagerman, Biochemistry **30**, 4268 (1991).

[204] Y. H. Wang, M. T. Howard, and J. D. Griffith, Biochemistry **30**, 5443 (1991).

[205] M. C. Wahl, C. Ban, C. Sekharudu, B. Ramakrishnan, and M. Sundaralingam, Acta Crystallogr. D **52**, 655 (1996).

[206] Z. Wu, F. Delaglio, N. Tjandra, V. B. Zhurkin, and A. Bax, Journal of Biomolecular NMR **26**, 297 (2003).

[207] A. Barbič, D. P. Zimmer, and D. M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **100**, 2369 (2003).

[208] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenář, and J. Feigon, Proc. Natl. Acad. Sci. U.S.A. **101**, 1177 (2004).

[209] D. Strahs and T. Schlick, J. Mol. Biol. **301**, 643 (2000).

[210] I. Fierro-Monti and M. B. Mathews, Trends Biochem. Sci. **25**, 241 (2000).

[211] A. Gebhardt, B. T. Laudenbach, and A. Pichlmair, J. Interferon Cytokine Res. **37**, 184 (2017).

[212] M. Ha and V. N. Kim, Nat. Rev. Mol. Cell Biol. **15**, 509 EP (2014).

[213] L. R. Saunders and G. N. Barber, FASEB J. **17**, 961 (2003).

[214] S.-H. Bae, H.-K. Cheong, J.-H. Lee, C. Cheong, M. Kainosho, and B.-S. Choi, Proc. Natl. Acad. Sci. U.S.A. **98**, 10602 (2001).

[215] J. Donovan, M. Dufner, and A. Korennykh, Proc. Natl. Acad. Sci. U.S.A. **110**, 1652 (2013).

[216] E. Uchikawa, M. Lethier, H. Malet, J. Brunel, D. Gerlier, and S. Cusack, Mol. Cell **62**, 586 (2016).

[217] R. Acevedo, D. Evans, K. A. Penrod, and S. A. Showalter, Biophys. J. **110**, 2610 (2016).

[218] R. Stefl, F. C. Oberstrass, J. L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R. B. Emeson, and F. H. T. Allain, Cell **143**, 225 (2010).

[219] I. Brukner, S. Susic, M. Dlakic, A. Savic, and S. Pongor, J. Mol. Biol **236**, 26 (1994).

[220] D. S. Goodsell, M. L. Kopka, D. Cascio, and R. E. Dickerson, Proc. Natl. Acad. Sci. U.S.A. **90**, 2930 (1993).